

Result No.	Score	Query 8		Length	DB	ID	Description
		Match					
1	102.5	7.4	1162	2	US-08-728-323A-2		Sequence 2, Appli
2	97	7.0	457	4	US-08-821-818-1		Sequence 1, Appli
3	97	7.0	457	4	US-08-821-818-5		Sequence 5, Appli
4	95	6.9	657	3	US-08-893-852A-3		Sequence 3, Appli
5	95	6.9	657	4	US-08-821-818-3		Sequence 3, Appli
6	92	6.7	441	4	US-08-244-603A-1		Sequence 1, Appli
7	90.5	6.6	221	4	US-09-069-023-22		Sequence 22, Appl
8	89.5	6.5	674	3	US-08-893-852A-1		Sequence 1, Appli
9	89	6.4	351	1	US-08-159-969-2		Sequence 2, Appli
10	89	6.4	352	2	US-08-726-306A-17		Sequence 17, Appl
11	89	6.4	391	2	US-08-244-951A-10		Sequence 10, Appl
12	89	6.4	391	2	US-08-389-011-23		Sequence 23, Appl

TELEPHONE: 212-278-0400

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted and is derived by analysis of the total score distribution.

```

; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-728-323A-2
Alignment Scores:
Pred. No.: 0.0194 Length: 1162
Score: 102.50 Matches: 36
Percent Similarity: 36.08% Conservative: 21
Best Local Similarity: 22.78% Mismatches: 59
Query Match: 7.43% Indels: 42
DB: 4 Gaps: 4

US-09-658-824-808 (1-781) x US-08-728-323A-2 (1-1162)
Qy 261 AGGCCCGAGGAGTGAGGGGCTCAGCGAAGCTGGGGTCTCTGGGGTATCCAGTCC 320
Db 422 LysylsGluAspGluGluAspGlyGlyAsnLysThrLeuSerIleGlnSer 441
Qy 321 CAGAAGCACCTGGAACCC-----CGACAGAA 347
Db 442 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 461
Qy 348 ATTCTGACTCCCGACAGCGGACAGGAGGCGCATGAGCGACACACACAAACACA 407
Db 462 ProLeuGlnGluProGln-----GlnGln 469
Qy 408 GAACACACAGCCAGTCC-----CAGAGCCCGAGTAATGAGAG 446
Db 470 GluProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 489
Qy 447 CCCCCAAAAGAAAGAACAGCAGCTGAAAGTCGGGATCCTACACCTGGGCGAGCAGACAGAA 506
Db 490 ProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Qy 507 GAACATCAGATACAGCTGAGATCCAGCTGCGGACATGGAAGTGATGTCGAAGAGCTG 566
Db 510 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 529
Qy 567 CATCAGTCAACACCGGGGATAAATCTGGATTGGTTCGGCGCTCAAGGTCAAGATAAT 626
Db 530 GlnGln-GlnGluPro-----GlnGlnGlnGln 538
Qy 627 ACCTAAGAGACACACTGTAAATGCCAAGCAGGTGAAGACCAACAA 678
Db 538 uProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 555

RESULT 2
US-08-821-818-1
; Sequence 1, Application US/08821818
; Patent No. 6146877
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
; TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/821,818
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Chan, Albert Wai-Kit
; REGISTRATION NUMBER: 36,479
; REFERENCE/DOCKET NUMBER: 51523
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-821-818-1
Alignment Scores:
Pred. No.: 0.0509 Length: 457
Score: 97.00 Matches: 48
Percent Similarity: 37.99% Conservative: 20
Best Local Similarity: 26.82% Mismatches: 67
Query Match: 7.03% Indels: 44
DB: 4 Gaps: 9

US-09-658-824-808 (1-781) x US-08-821-818-1 (1-457)
Qy 112 AGTCAGAGTCAAGAGGACCAAGAGGCTCCGAGGGTGTGTGTCAGTCACTACAGT 171
Db 220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgValTrpGluTyrHisThrArg 239
Qy 172 GAGAAGGCCCTCGAAGTCGTCGCTCTCATCGGTGCACGCCCATCGACCTTCTTCTGT 231
Db 240 GluArgProLysGlnGlnGluGlyGluThrLysProGluGlnHisArgAlaGlnSerHis 259
Qy 232 CTCCTCAGGCCATACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 279
Db 260 ProCysGlnAsnAlaGluAlaGluGluGlyGlyProGluThrSerValCysSerGlySer 279
Qy 280 GCTCAGCGAAGCTGGGGTCTGTGGGGTATCCAGTCCAGAGACACCTGGAAACCCC 339
Db 280 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 292
Qy 340 GACAGAAGATTCGAGTCCCTCCCA-----GACGGGACCAAGAGAGAGGAGGAGGAGGAGGAG 393
Db 292 pThrGluGluGlu-GluAspSerAspLeuAspSerAlaGluGluAspThrAla----- 309
Qy 394 CACACACAAACACAGAACACACACAGC----- 419
Db 310 --HisThrCysThrThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 329
Qy 420 -----CAGTCCCGAGGCCAGTAATGGAGAG---CCCCAAAAGAAAGAACACGAGCTG 471
Db 329 lyGluAspThrGluGluGluAspGlyAspTrpAspSerAlaGluGluAspAlaSerG 349
Qy 472 AAAGT---CGGGATCCT-----ACACCTGGGC 495
Db 349 InSerCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 369
Qy 496 AGCAGACAGAAAGATCAGATACAGCTGAGATCCAGTCCGAGGAGGAGGAGGAGGAGGAGGAG 548
Db 369 luAspThrGluGluGluAspSerGluAsnValAlaProValAspSerGlu 386

RESULT 3
US-08-821-818-5
; Sequence 5, Application US/08821818
; Patent No. 6146877
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
; TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF
```

```

Db      349  lnSerCystThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 549
Qy      496  ACAGACAGACAGAAGAAGACAGGATACAGCTCAGATCCAGTCGCAGCATGGAA 548
        :  |||||:::  ||:::  :::  |||  |||
Db      369  luAspThrGlUGluGluAspSerGluAsnValAlaProValAspSerGlu 386
        :  |||||:::  ||:::  :::  |||  |||

RESULT 4
US-08-893-852A-3
; Sequence 3, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A

```

: PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER:  
 : FILING DATE:  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Billings, Lucy J.  
 : REGISTRATION NUMBER: 36,749  
 : REFERENCE/DOCKET NUMBER: PF-0341 US  
 : TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

```

; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 657 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     LIBRARY: GenBank
;     CLONE: 53041
;   US-08-893-852A-3

Alignment Scores:
Pred. No.: 0.0963
Score: 95.00
Percent Similarity: 35.77%
Best Local Similarity: 24.80%
Query Match: 6.88%
DB: 3

US-09-658-824-808 (1-781) x US-08-893-852A-3 (1-657)

Qy 55 GCTACTGACACGCGGGTAGTGCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGT 114
||||||| ||: ::||| |||
Db 212 AlaThrGluGluYsThrGluAsnLysAlaAspProSerAsnProSer----- 229
||||||| |||

Qy 115 GAGAGTGAAGAGAACACAGCAGCGCTCCGGAGGGTTGTTGGTCAGTGCATCAGACTGAG 174
|||||
Db 230 -----SerGlySerHisSerArgAlaTrpGluTyrTyrSerArgGlu 243
|||||

Qy 175 AAGGCGCTCGAAGTCGTGTCCTCTCTACGCGGTGCCACGCCCATGCACCTTCTTGTCTC 234
|||

```





US-08-244-603A-1  
: Sequence 1, Application US/08244603A  
: Patent No. 6200768  
: GENERAL INFORMATION:  
: APPLICANT: Mandelkow, Eva-Maria  
: APPLICANT: Mandelkow, Eckhard  
: APPLICANT: Lichtenberg-Kraag, Birgit  
: APPLICANT: Biernat, Jacek  
: APPLICANT: Drewes, Gerard  
: APPLICANT: Steiner, Barbara  
: TITLE OF INVENTION: No. 6200768el Tools For The Diagnosis And  
: TITLE OF INVENTION: Treatment Of Alzheimer's Disease  
: NUMBER OF SEQUENCES: 1  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
: ADDRESSEE: Borun  
: STREET: 233 South Waker Drive, 6300 Sears Tower  
: CITY: Chicago  
: STATE: Illinois  
: COUNTRY: United States of America  
: ZIP: 60606-6402  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Tape  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/244,603A  
: FILING DATE:  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Joseph A. Williams, Jr.  
: REGISTRATION NUMBER: 38,659  
: REFERENCE/DOCKET NUMBER: 28384/32778  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 312-474-6300  
: TELEFAX: 312-484-0448  
: TELEX: 25-3856  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 441 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-244-603A-1

Alignment Scores:  
Pred. No.: 0.171 Length: 441  
Score: 92.00 Matches: 59  
Percent Similarity: 39.90% Conservative: 24  
Best Local Similarity: 28.37% Mismatches: 80  
Query Match: 6.67% Indels: 45  
DB: 11 Gaps: 11

US-09-658-824-808 (1-781) x US-08-244-603A-1 (1-441)  
QY 76 GGTCCACAGCAGATCCAACTGGAGTTGAAGTGTGAGTGAGAGGAGCAACGCA 135  
Db 100 GlyThrThrAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAla 119  
QY 136 GGCTTCGGAGGGTTGTGTGCTAGTCACTCAGTACAGAGGCGCTCGAAGTCTGCTC 195  
Db 120 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 133  
QY 196 CCTCTCATGCGTGCACGCCCATGGACCTCTTCTGTCGTCACGGCCATACTAGGAG 255  
Db 133 p-----GlyThrGlySerAsp-----As 139  
QY 256 GAGGAGGGCCGAGGAGTGGAGGGCTCAGCGCA----- 289  
Db 139 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 159

QY 290 -----AGCTGGGTGCTCTTGGGGGTATTCGGAGTC-----CCAGAGCACC 333  
Db 159 oProGlyGlnLysGlnAlaAsnAlaThrArgIleProAlaLysThrProAlaApr 179  
QY 331 TCGAACCCCGACAGACAGATTCTGGACTCCCGACAGCGGACGAGAGGCGCATCAG 390  
Db 179 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 198  
QY 391 CGACACACACAAACACA---GAACACACACAGCCAGTCCCGAGGACCCAGTAATGGAGGC 447  
Db 198 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 218  
QY 448 CCAAAAAGCAAGAACCA-----GCAGCTGAAAGTCGGGATCTCTACACCTCGGCAGCAGA 501  
Db 218 roProThrArgGluProLysLysValAlaValAlaArgThrProProLysSerProSerS 238  
QY 502 CAGAACAAGATCAG---GATACAGCTGAGATCCAGTGCCTGGAGATCGAAGGTGATCTGC 558  
Db 238 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 256  
QY 559 AAGAGCTGCATCAGTCAACACACCGGGGATAAATCTGCGATTGGCTTCGGCGTCAAGGTG 618  
Db 257 LysSerLysIleGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly---LysVal 275  
QY 619 AAGATAATACCTAAAGAG 636  
Db 276 GlnIleIleAsnLysLys 281

RESULT 7  
US-09-069-023-22  
: Sequence 22, Application US/09069023A  
: Patent No. 6348573  
: GENERAL INFORMATION:  
: APPLICANT: Nunez, Gabriel  
: APPLICANT: Inohara, Naohiro  
: TITLE OF INVENTION: Koseki, Takeyoshi  
: TITLE OF INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS  
: FILE REFERENCE: UM-03333  
: CURRENT APPLICATION NUMBER: US/09/069,023A  
: CURRENT FILING DATE: 1998-04-27  
: NUMBER OF SEQ ID NOS: 38  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 22  
: LENGTH: 221  
: TYPE: PRT  
: ORGANISM: Mus musculus  
: US-09-069-023-22

Alignment Scores:  
Pred. No.: 0.186 Length: 221  
Score: 90.50 Matches: 36  
Percent Similarity: 37.32% Conservative: 17  
Best Local Similarity: 25.35% Mismatches: 47  
Query Match: 6.56% Indels: 42  
DB: 4 Gaps: 6

US-09-658-824-808 (1-781) x US-09-069-023-22 (1-221)  
QY 293 TGGGGTGTGTGGG-----GGTATC 313  
Db 93 TrpGlnHisValGlyProGlyTyrArgAspArgSerTyrAspProProCysProGlyHis 112  
QY 314 CGAGTCCCAAGAGCAGCTGGAACCCGACAGAGATTCTGGATCTCCCGACGCGGACAG 373  
Db 113 TrpThrProGluAlaProSerSerGlyThrThrCysProGlyLeuProArgAlaSerGlu 132  
QY 374 GAGAGGGAC-----GGCATGAGCGACACACACAAACACAGACCCAGTCCCA 427  
Db 133 GluG IGuIleGlyGlyProGluAspSerGluAla-ValGlnProArgThrProGluG 152  
QY 428 G-----GAGCCCGAGTAATGGA-----GA 445

US-08-244-603A-1  
: Sequence 1, Application US/08244603A  
: Patent No. 6200768  
: GENERAL INFORMATION:  
: APPLICANT: Mandelkow, Eva-Maria  
: APPLICANT: Mandelkow, Eckhard  
: APPLICANT: Lichtenberg-Kraag, Birgit  
: APPLICANT: Biernat, Jacek  
: APPLICANT: Drewes, Gerard  
: APPLICANT: Steiner, Barbara  
: TITLE OF INVENTION: No. 6200768el Tools For The Diagnosis And  
: TITLE OF INVENTION: Treatment Of Alzheimer's Disease  
: NUMBER OF SEQUENCES: 1  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
: ADDRESSEE: Borun  
: STREET: 233 South Waker Drive, 6300 Sears Tower  
: CITY: Chicago  
: STATE: Illinois  
: COUNTRY: United States of America  
: ZIP: 60606-6402  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Tape  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/244,603A  
: FILING DATE:  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Joseph A. Williams, Jr.  
: REGISTRATION NUMBER: 38,659  
: REFERENCE/DOCKET NUMBER: 28384/32778  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 312-474-6300  
: TELEFAX: 312-484-0448  
: TELEX: 25-3856  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 441 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-244-603A-1

Alignment Scores:  
Pred. No.: 0.171 Length: 441  
Score: 92.00 Matches: 59  
Percent Similarity: 39.90% Conservative: 24  
Best Local Similarity: 28.37% Mismatches: 80  
Query Match: 6.67% Indels: 45  
DB: 11 Gaps: 11

US-09-658-824-808 (1-781) x US-08-244-603A-1 (1-441)  
QY 76 GGTCCACAGCAGATCCAACTGGAGTTGAAGTGTGAGTGAGAGGAGCAACGCA 135  
Db 100 GlyThrThrAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAla 119  
QY 136 GGCTTCGGAGGGTTGTGTGCTAGTCACTCAGTACAGTACAGAGGCCCTCGAAGTCGTC 195  
Db 120 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 133  
QY 196 CCTCTCATGCGTGCACGCCCATGGACCTCTTCTGTCGTCACGGCCATACTAGGAG 255  
Db 133 p-----GlyThrGlySerAsp-----As 139  
QY 256 GAGGAGGGCCGAGGAGTGGAGGGCTCAGCGCA----- 289  
Db 139 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 159

QY 290 -----AGCTGGGTGCTCTTGGGGGTATTCGGAGTC-----CCAGAGCACC 333  
Db 159 oProGlyGlnLysGlnAlaAsnAlaThrArgIleProAlaLysThrProAlaApr 179  
QY 331 TCGAACCCCGACAGACAGATTCTGGACTCCCGACAGCGGACGAGAGGCGCATCAG 390  
Db 179 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 198  
QY 391 CGACACACACAAACACA---GAACACACACAGCCAGTCCCGAGGACCCAGTAATGGAGGC 447  
Db 198 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 218  
QY 448 CCAAAAAGCAAGAACCA-----GCAGCTGAAAGTCGGGATCTCTACACCTCGGCAGCAGA 501  
Db 218 roProThrArgGluProLysLysValAlaValAlaArgThrProProLysSerProSerS 238  
QY 502 CAGAACAAGATCAG---GATACAGCTGAGATCCAGTGCCTGGAGATCGAAGGTGATGTC 558  
Db 238 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 256  
QY 559 AAGAGCTGCATCAGTCAACACACCGGGGATAAATCTGCGATTGGCTTCGGCGTCAAGGTG 618  
Db 257 LysSerLysIleGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly---LysVal 275  
QY 619 AAGATAATACCTAAAGAG 636  
Db 276 GlnIleAsnLysLys 281

RESULT 7  
US-09-069-023-22  
: Sequence 22, Application US/09069023A  
: Patent No. 6348573  
: GENERAL INFORMATION:  
: APPLICANT: Nunez, Gabriel  
: APPLICANT: Inohara, Naohiro  
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS  
: TITLE OF INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS  
: FILE REFERENCE: UM-03333  
: CURRENT APPLICATION NUMBER: US/09/069,023A  
: CURRENT FILING DATE: 1998-04-27  
: NUMBER OF SEQ ID NOS: 38  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 22  
: LENGTH: 221  
: TYPE: PRT  
: ORGANISM: Mus musculus  
: US-09-069-023-22

Alignment Scores:  
Pred. No.: 0.186 Length: 221  
Score: 90.50 Matches: 36  
Percent Similarity: 37.32% Conservative: 17  
Best Local Similarity: 25.35% Mismatches: 47  
Query Match: 6.56% Indels: 42  
DB: 4 Gaps: 6

US-09-658-824-808 (1-781) x US-09-069-023-22 (1-221)  
QY 293 TGGGGTGTGTGGG-----GGTATC 313  
Db 93 TrpGlnHisValGlyProGlyTyrArgAspArgSerTyrAspProProCysProGlyHis 112  
QY 314 CGAGTCCCAAGAGCAGCTGGAACCCGACAGAGATTCTGGATCTCCCGACGCGGACAG 373  
Db 113 TrpThrProGluAlaProSerSerGlyThrThrCysProGlyLeuProArgAlaSerGlu 132  
QY 374 GAGAGGGAC-----GGCATGAGCGACACACACAAACACAGACCCAGTCGCCCA 427  
Db 133 GluG IGuIleGlyGlyProGluAspSerGluAla-ValGlnProArgThrProGluG 152  
QY 428 G-----GAGCCCGAGTAATGGA-----GA 445  
Db 428 G-----GAGCCCGAGTAATGGA-----GA 445

152	uProGluLeuGluAlaGluAlaThrLysGlyAspGluProAspLeuGluGlnGluMetGI	172
Db		
446	GCCCAAAAGAGAACCAGCAGCTGAAGTCGGGATCCTACACTGGCAGCAGACAGA	505
Qy	:::     ::        ::        ::        ::        ::        ::	
172	uProGluProGluProGluValCluProGluProGluProGluProGluProGluProGI	192
Db		
506	ACAAGATCAGGATCAGCTCAGATCCCAGTGCAGCATGGAAGGTGATCTGCAAGAGCT	565
Qy	:::     ::        ::        ::        ::        ::        ::	
192	uProGluProGluProGluProGluProGluProGluArgGluProAspPheGlnGlu--	211
Db		
566	GCATCAGTCAAACACCGGGGATAATCTCGATTGTTCCGGCGCTCAGGTGTAAGATAA	625
Qy	:::     ::        ::        ::        ::        ::        ::	
212	-----GlvAspGluSer-----GluGlycvsGluas	220
Db		

```

: RESULT 8
: US-08-893-852A-1
: Sequence 1, Application US/08893852A
: Patent No. 6080538
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Lal, Preeti
: APPLICANT: Shah, Purvi
: APPLICANT: Corley, Neil C.
: TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.

```

US-09-658-824-808 (1-781) x US-08-893-852A-1 (1-674)

Qy	186	AGTCGTGTCCTCTCTATCGGGTGCACGCCCATGGACCTTCTTCTCTCGTCACGGCCAT	245
Db	249	SerProArgSerGlySerAspProArgSerTrpGluTyrArgSerGlyAlaSer	268
Qy	246	AACTAGGAGGAGGCGCGAGGAGTGGAGGGCTCAGCGCAAGCTGGG	296
Db	269	GluGluLysGluGluLysAlaHisGluGluThrGlyLysGlyGluAlaAlaProGlyPro	288
Qy	296	-----	296
Db	289	GlnSerSerAlaProAlaGlnArgProGlnLeuLysSerTrpTrpCysGlnProSerAsp	308
Qy	297	-----GTGCTGTTGGGGTATCCGAGTC	319
Db	309	GluGluGluSerGluValLysAlaLeuGlyAlaAlaGluLysAspGlyGluAlaGluCys	328
Qy	320	-----CCAGAAGCA	337
Db	329	ProProCysIleProProProSerAlaPheLeuLysAlaTrpValTrpProGlyGlu	348
Qy	338	CCGACACAGAATCTGCACCTCCACAGCGGCACGACGAGGACGCCATGACGCACACA	397
Db	349	AspThrGluGluGluGluAspGluGluGluAspGluAspSerAspGlySerAspGlu	368
Qy	398	CACAAACACACACACACAGCAGCTCCACGAGGCCCATATGGA	443
Db	369	GluGlu-GlyGluAlaGluAlaSerSerThrProAlaThrGlyValPheLeuLysSer	388
Qy	443	-----	443
Db	388	rTrpValTrpGlnProGlyGluAspThrGluGluGluGluAspGluAspSerAspThrGln	408
Qy	444	-----CAGCCCCAAAAGAAGAACCCAGCAGCTGAAGTCGGGATCCTACA	488
Db	408	ySerAlaGluAspGluArgGluAlaGluThrSerAlaSerThrProProAlaSerAlaPhe	428
Qy	489	-----CCTGGCGCACACAGAGAAGAT	512
Db	428	eLeuLysAlaTrpValTyrArgProGlyGluAspThrGluGluGluGluAspGluAspVal	448
Qy	513	-----CAGGATACAGCTGAGATCCCGAGTCGCGACATGGAAGGTGATCT	556
Db	448	lAspSerGluAspLysGluAspAspSerGluAlaAlaLeuGlyGluAlaGluSerAspPr	468
Qy	557	GCAAGAGCTGCATCAGTCAACACCGGGGTAATCTGGATTTCGGTCCGGCGTCAAGG	616
Db	468	oHisProSerHisProAspGlnSerAlaHisPheArgGlyTyrP61yTyr--ArgProGln	487
Qy	617	TGAAGATAATACCTAAAGAGGAGAACACTGTAAAATGCCAGAGCAGGTGAAGACAACCA	675
Db	487	ylsGluThr-GluGluGluGluAlaAlaGlu-----AspTrpGluGluAlaGluPro	504

```

RESULT 9
US-08-159-969-2
; Sequence 2, Application US/08159969
; Patent No. 5492812
;
GENERAL INFORMATION:
;
APPLICANT: Voorchels, Paul H.
;
TITLE OF INVENTION: Diagnostic Method for Alzheimer's
;
TITLE OF INVENTION: Disease
;
NUMBER OF SEQUENCES: 2
;
CORRESPONDENCE ADDRESS:
;
ADDRESSEE: Pennie & Edmonds
;
STREET: 1155 Avenue of the Americas
;
CITY: New York
;
STATE: New York
;
COUNTRY: U.S.A.
;
ZIP: 10036-2711
;
COMPUTER READABLE FORM:
;
MEDIUM TYPE: Floppy disk
;
COMPUTER: IBM PC compatible
;

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/159,969  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/738,778  
 FILING DATE: 01-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 4697-040  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212 869-8864/9741  
 TELEFAX: 212 869-8864/9741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 351 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-159-969-2

Alignment Scores:  
 Pred. No.: 0.325 Length: 351  
 Score: 89.00 Matches: 58  
 Percent Similarity: 40.10% Conservative: 25  
 Best Local Similarity: 28.02% Mismatches: 79  
 Query Match: 6.45% Indels: 45  
 DB: 1 Gaps: 11

US-09-658-824-808 (1-781) x US-08-159-969-2 (1-351)  
 QY 76 GCTCCACAGCAGATCCAACTGGAGTTGAAGTGTGAGTGAAGAGGAAACACGACA 135  
 Db 42 GlyLeuLysAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAlaA 61  
 QY 136 GCCTTCGGAGGGTTGTGTGCTCACTCAGACTGAGAGGCCCTCGAAGTCGTCGTC 195  
 Db 62 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 75  
 QY 196 CCTCTACGCGTCCAGGCCATCTCTGCTCGTCACGGCCATACTAGGAG 255  
 Db 75 p-----GlyThrGlySerAsp-----As 81  
 QY 256 GAAGGAGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 289  
 Db 81 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 101  
 QY 290 -----AGCTGGGGTCTGTGGGGGATCCGAGTC-----CCAGAAGCACC 330  
 Db 101 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProAlaPr 121  
 QY 331 TGGAAACCCGACACAAAGATCTTGACATCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 390  
 Db 121 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 140  
 QY 391 CGACACACACAAACACA---GAACCCACACACAGCTCCCGAGGAGGCCAGTAATGGAGAGC 447  
 Db 140 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 160  
 QY 448 CCAAAAAGAAGACCA-----CGAGCTGAAGTCGGGATCCCTACACTGGGCGACGAGA 501  
 Db 160 roProThrArgGluProLysLysValAlaValAlaArgThrProProLysSerProSers 180  
 QY 502 CAGAAGAAGATCAG---GATACAGCTGAGATCCCGAGTCCGCGACATGAAGGTGATCTGC 558  
 Db 180 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 198  
 QY 559 AAGAGCTGCATCAGTCAACACCGGGGATAAATCTGGATTGTTGGCGGTCGCAAGTG 618

Db 199 LysSerLysIleGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly---LysVal 211  
 QY 619 AAGATAATACCTAAA 633  
 Db 218 GlnIleValTyrLys 222

RESULT 10  
 US-08-726-306A-17  
 ; Sequence 17, Application US/08726306A  
 ; Patent No. 5958684  
 ; GENERAL INFORMATION:  
 ; APPLICANT: van Leeuwen, Frederik Willem  
 ; APPLICANT: Burbach, Johannes Peter Henri  
 ; APPLICANT: Grosveld, Franklin G.  
 ; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS  
 ; NUMBER OF SEQUENCES: 189  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Banner & Witcoff, Ltd.  
 ; STREET: 1 Financial Center  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: US  
 ; ZIP: 02111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WordPerfect 6.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/726,306A  
 ; FILING DATE: 02-Oct-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 95/20080.4  
 ; FILING DATE: 02-Oct-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/009,832  
 ; FILING DATE: 01-Jan-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Williams Ph.D. Kathleen M.  
 ; REGISTRATION NUMBER: 34,380  
 ; REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 345-9100  
 ; TELEFAX: (617) 345-9111  
 ; INFORMATION FOR SEQ ID NO: 17:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 352 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-726-306A-17

Alignment Scores:  
 Pred. No.: 0.325 Length: 352  
 Score: 89.00 Matches: 58  
 Percent Similarity: 40.10% Conservative: 25  
 Best Local Similarity: 28.02% Mismatches: 79  
 Query Match: 6.45% Indels: 45  
 DB: 2 Gaps: 11

US-09-658-824-808 (1-781) x US-08-726-306A-17 (1-352)  
 QY 76 GGTCCACAGCAGATCCAACTGGGAGTTGAAGTGTGAGTGAAGAGGAAACACGACA 135  
 Db 42 GlyLeuLysAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAlaA 61  
 QY 136 GCCTTCGGAGGGTTGTGTGCTCAGTCACTCAGAGTGAAGGCCCTCGAAGTCGTCGTC 195  
 Db 62 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 75  
 QY 196 CTTCTCATCGGTGCGGATCCAGGCCATCTCTGCTCGTCACGGCCATACTAGGAG 255

[illegible]

```
Db 75 p-----GlyThrGlySerAsp-----As 81
Qy 256 GAAGGAGGCGCAGAGTGGAGGGCTCAGCGCA----- 289
Db 81 pLysLysAlaLysGlyGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 101
Qy 290 -----AGCTGGGTGCTGTGGGGTATCCGAGTC-----CCAGAAGCACC 330
Db 101 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProAlaPr 121
Qy 331 TGGACCCCGACAGAGATTCTGCACCTCCCGAGCGGCGCAGAGAGGCGCATGAG 390
Db 121 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 140
Qy 391 CGACACACAAACACACA---GAACACACACAGCCAGTCCAGCGAGCCAGTAATGGAGGC 447
Db 140 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 160
Qy 448 CCCAAAGAAGACACCA-----CGAGCTGAAGTCGGGATCTCAGCTCGGCGTCAAGGTG 618
Db 160 roProThrArgGluProLysLysValAlaValAlaArgThrProProLysSerProSerS 180
Qy 502 CAGAGAAGATCAG---GATACAGCTGAGATCCAGTCGCGCATGGAAGGTGATCTGC 558
Db 180 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 198
Qy 559 AAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGGATTGGGTTCGGCGTCAAGGTG 618
Db 199 LysSerLysIleGlySerThrGluAsnLeuLysHisGlnProGlyGlyLys---LysVal 217
Qy 619 AAGATAATACCTAAA 633
Db 218 GlnIleValTyrLys 222

RESULT 11
US-08-244-951A-10
; Sequence 10, Application US/08244951A
; Patent No. 5843779
; GENERAL INFORMATION:
; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;
; APPLICANT: VANDERMEEREN, EUGEN; VAN DE VOORDE, ANDRE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: DIRECTED AGAINST THE MICROBUTULE-ASSOCIATED
; TITLE OF INVENTION: PROTEIN TAU, HYBRIDOMAS SECRETING THESE
; TITLE OF INVENTION: ANTIBODIES, ANTIGEN RECOGNITION BY THESE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AND THEIR APPLICATIONS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN & MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244, 951A
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/03499
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP/92/403403.6
; FILING DATE: 14-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 410.003A
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391
; TYPE: Amino Acid
; STRANDEDNESS: Unknown
; TOPOLOGY: Unknown
; FEATURE:
; NAME/KEY: mTHFMPH-taul fusion protein
; US-08-244-951A-10

Alignment Scores:
Pred. No.: 0.339 Length: 391
Score: 89.00 Matches: 58
Percent Similarity: 40.10% Conservative: 25
Best Local Similarity: 28.02% Mismatches: 79
Query Match: 6.45% Indels: 45
DB: 2 Gaps: 11

US-09-658-824-808 (1-781) x US-08-244-951A-10 (1-391)
Qy 76 GGTCCACAGCGAGATCCCACTGGAGTGAAGTGTGAGTGAGAGGAACACCA 135
Db 81 GlyLeuLysAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAla 100
Qy 136 GCGTTCGGGAGGGTGTGTGCTCAGTGACTCAGAGTGAGAAAGGCCCTCGAAGTCTGCTC 195
Db 101 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 114
Qy 196 CCTCTCATGGCGTGCACGCCCATGGAGCTTCTTGTCTCAGCGGCATAACTAGGAG 255
Db 114 p-----GlyThrGlySerAsp-----As 120
Qy 256 GAAGGAGGCGCAGAGTGGAGGGCTCAGCGCA----- 289
Db 120 pLysLysAlaLysGlyLysAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 140
Qy 290 -----AGCTGGGTGCTGTGGGGTATCCGAGTC-----CCAGAAGCACC 330
Db 140 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProAlaPr 160
Qy 331 TGGAAACCCCGACAGAGATTCTGCACCTCCCGAGCGGCGCAGAGAGGCGCATCAG 390
Db 160 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 179
Qy 391 CGACACACACAAACACACA---GAACACACACAGCCAGTCCCGAGCGCCAGTAATGGAGGC 447
Db 179 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 199
Qy 448 CCCAAAGAAGACACCA-----GCAGCTGAAAGTCGGGATCTCAGCTCGGCGCAGCA 501
Db 199 roProThrArgGluProLysLysValAlaValAlaArgThrProProLysSerProSerS 219
Qy 502 CAGAAGAAGATCAG---GATACAGCTGAGATCCCGAGTCGCGCATGGAAGGTGATCTGC 558
Db 219 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 237
Qy 559 AAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGGATTGGGTTCGGCGTCAAGGTG 618
Db 238 LysSerLysIleGlySerThrGluAsnLeuLysHisGlnProGlyGlyLys---LysVal 256
Qy 619 AAGATAATACCTAAA 633
Db 257 GlnIleValTyrLys 261

RESULT 12
US-08-389-011-23
; Sequence 23, Application US/08389011
; Patent No. 5861257
; GENERAL INFORMATION:
; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;
```



RESULT 14

```

QY 196 CCTCTCATGCGGCGCCAGCCCATGACCTTCTTCTCTCCTCAGCGCCATACTAGGGAG 255
Db 114 p-----GlyThrGlySerAsp-----As 120
QY 256 GAAGGAGGCGCAGGAGCTCAGGGCTCAGCGCA-----289
Db 120 pLysLysAlaLysGlyAlaLysGlyLysThrLysLysIleAlaThrProArgGlyAlaLysPr 140
QY 290 -----ACGTGGGGTCTCTTGGGGGTATCCGAGTC-----CCAGAAGCACC 330
Db 140 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProProAlaLysPr 160
QY 331 TGGNACCCCGACAGAAATCTCGACTCCCGACGCGGACGAGGAGCGGCATGAG 390
Db 160 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 179
QY 391 CGACACACACAAACACA---GAACACACACAGCCAGTCCCGAGGAGCCCGAATGGAGAGC 447
Db 179 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 199
QY 448 CCAAAAAGAGAACCA-----GCAGCTGAAAGTCGGGATCCTACACCTGGCGACGAGA 501
Db 199 roProThrArgGluProLysLysValAlaValAlaValArgThrProProLysSerProSerS 219
QY 502 CAGAAGAAGATCAG---GATACAGCTGAGATCCCGAGTCGCGACATGGAAGGTGATCTGC 558
Db 219 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 237
QY 559 AAGAGCTGCATCAGTCAAAACACCGGGGATTAATCTGGATTGGTTCCCGCGCTCAAGGTG 618
Db 238 LysSerLysLysGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly---LysVal 256
QY 619 AAGATAATACCTAAA 633
Db 257 GlnIleValYrLys 261

RESULT 15
US-08-447-591-2
; Sequence 2, Application US/08447591
; Patent No. 5591440
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLACE, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,591
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/059,031
; FILING DATE: 07-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5347.US.01
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-447-591-2

Alignment Scores:
Pred. No.: 0.348 Length: 228
Score: 88.00 Matches: 45
Percent Similarity: 35.79% Conservative: 23
Best Local Similarity: 23.68% Mismatches: 59
Query Match: 6.17% Indels: 63
DB: 1 Gaps: 9

US-09-658-824-808 (1-781) x US-08-447-591-2 (1-228)

QY 744 GATAGTTTAAGTCAATATCTAATATAAAACACGCTTTCGCTTGTGTTTTCAGCTTGT----- 691
Db 52 AsnSerGlnSerGlnIleSerHisSerProThrCysCysProProIleCysProGly 71
QY 690 -----CTTCATTTAAACCTTGTGGTTGCTCTTACCTGCTCTTCTGCGATTTTACAGTG 640
Db 72 TyrArgTrpMetCysLeuArgPheIleIlePheLeuCysIle-----LeuLeuLeu 89
QY 639 TTCTCTTTTAGGTATTATCTTACCTTACGCGCGGACCCAAATCCAGATTTATCCCGG 580
Db 90 CysLeuIlePheLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysPro--- 108
QY 579 TGTTTGACTGATGACGCTTTCAGATCACCTTCCATGTGCG-----538
Db 109 LeuIleProGlySerSerThrThrSerThrGlyProCysArgAsnThrThrCysThrThr 128
QY 537 ---GCACCTGGGATCTCAGCTGATCTCTCTCTGCTGCTGCCAGGTGTAGG--- 483
Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148
QY 482 -----ATCCCGACTTTCAGCTGCTGTTCTTTC-----456
Db 148 nCysThrCysIleProIleProSerSerTrpAlaPheAlaLysPheLeuTrpGluTrpAl 168
QY 455 -----TTTTTGGGCT 445
Db 168 aSerValArgPheSerTrpLeuSerLeuLeuValProPheValGlnTrpPheValGlyLe 188
QY 444 CTCCATTACTGGGCTCCTGGGACTG-----GCTGTGTGGTTCTGTGTTTGTGTGTGTCG 391
Db 188 uSerProThrValTrpLeuSerValIleTrpMetMetTrpTyr-----202
QY 390 CTCATGCGGCTCCTCTCTGCTGCCGTCGGGA-----GTCCAGAAATCTTCTGTCGG 337
Db 203 -----TrpGlyProSerLeuTyrAsnIleLeuSerPr 213
QY 336 GTTCCAGTGCTTCTGGGACTCGGATC 309
Db 213 oPheMetProLeuLeuProIlePheTyr 222

RESULT 16
US-08-450-943-2
; Sequence 2, Application US/08450943
; Patent No. 5593825
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLIS, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION

```

```

Db 188 userProthrValtrpLeuSerValIleTrpMetMetTrpTyr----- 202
Qy 390 CTCAATGCCGTCCTCTCTGCTGCCGTGGGA-----GTCCAGAAATCTTGTGTCGG 337
Db 203 -----TTPGlyProSerLeuTyrAsnIleLeuSerPr 213
Qy 336 GTTCCAGGTGCTCTGGCACTCGATAC 309
Db 213 oPheMetProLeuLeuProfilePheTyr 222

RESULT 17
US-08-059-031-2
; Sequence 2, Application US/08059031
; Patent No. 595739
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLACE, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, HARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5347.US.01
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-059-031-2

Alignment Scores:
Pred. No.: 0.348 Length: 228
Score: 88.00 Matches: 45
Percent Similarity: 35.79% Conservative: 23
Best Local Similarity: 23.68% Mismatches: 59
Query Match: 6.17% Indels: 63
DB: 1 Gaps: 9

US-09-658-824-808 (1-781) x US-08-059-031-2 (1-228)
Qy 744 GATAGTTAAGTCAAAATCTTAATATAAAACGCTGCGTGTGTTTCAGCTTGT----- 691
Db 52 AsnSerGlnSerGlnIleSerSerHisSerProThrCysCysProPheCysProGly 71
Qy 690 -----CCTCATTTAAACTTGCGTGTCTCTTACCTGCTCTCGCATTTTACAGTG 640
Db 72 TyrArgTrpMetCysLeuArgPheIleIlePheLeuCysIle-----LeuLeuLeu 89
Qy 639 TTCCTCTTATAGGTATATCTTCACTTCACTTCAGCCCGGACCAAAATCCAGATTTATCCCGG 580

```



```

Db 90 CysLeuilePheLeuLeuValLeuAspTyrGlnGlyMetLeuProValCysPro--- 108
QY 579 TGTGTGACTGATCAGCTCTTGAGATCACCCTTCACGTGCGC----- 538
Db 109 LeuileProGlySerSerThrThrSerThrGlyProCysArgAsnThrCysThrThr 128
QY 537 ---GCACTGGGATCTCAGCTGTATCTCTGATCTTCTTCTGTCTGCTGCCAGGTGTAGG-- 483
Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148
QY 482 -----ATCCGACTTTCAGCTGCTGCTGCTTC----- 456
Db 148 nCysThrCysileProileProSerSerTrpAlaPheAlaLysPheLeuTrpGluTrpAl 168
QY 455 -----TTTTTGGGGCT 445
Db 168 aSerValArgPheSerTrpLeuSerLeuLeuValProPheValGlnTrpPheValGlyLe 188
QY 444 CTCCATTACTGGCTCTCTGGGACTG-----GCTGTGTGTTCTGTGTGTGTGTGTGTG 391
Db 188 userProThrValTrpLeuSerValileTrpMetMetTrpTyr----- 202
QY 390 CTCATGCGCTCCCTCTCTGCTGCTGCTGCGGA-----GTCCAGAAATCTTCTGTGCGG 337
Db 203 -----TrpGlyProSerLeuTyraAsnileLeuSerPr 213
QY 336 GTTCCAGGTGCTTCTGGGACTCGGATAC 309
Db 213 oPheMetProLeuLeuProilePheTyr 222

```

## RESULT 18

US-08-450-942-2

; Sequence 2, Application US/08450942

; Patent No. 5925512

; GENERAL INFORMATION:

; APPLICANT: CARMAN, WILLIAM

; APPLICANT: DECKER, RICHARD H

; APPLICANT: WALLACE, LESLEY

; APPLICANT: MIMMS, LARRY T

; APPLICANT: SOLOMON, LARRY R

; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D

; STREET: ONE ABBOTT PARK ROAD

; CITY: ABBOTT PARK

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08450.942

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: FOREMSKI, PRISCILLA E.

; REGISTRATION NUMBER: 33,207

; REFERENCE/DOCKET NUMBER: 5347.US.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 708-937-6365

; TELEFAX: 708-938-2623

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 228 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

;

;

;

;

;

;

;

;

;

US-08-450-942-2

Alignment Scores:

Pred. No.: 0.348

Score: 88.00

Percent Similarity: 35.79%

Best Local Similarity: 23.68%

Query Match: 6.17%

DB: 2

Length: 228

Matches: 45

Conservative: 23

Mismatches: 59

Indels: 63

Gaps: 9

US-09-658-824-808 (1-781) x US-08-450-942-2 (1-228)

QY 744 GATAGTTTAAGTCAATATCTAATATAAAACACGCTGCGTGTGTTTCAGCTTGT----- 691

Db 52 AsnSerGlnSerGlnileSerHisSerProThrCysCysProProileCysProGly 71

QY 690 -----CTTCATTTAAACTTGTGTTGCTTCTTACCTGCTTCTGCGCATTTTACAGTG 640

Db 72 TyrArgTrpMetCysLeuArgArgPheIleIlePheLeuCysile-----LeuLeuLeu 89

QY 639 TTCCTCTTTAGGTATTATCTTACCTTGGCGCGGAAACCCAAATCCAGATTATATCCCGCG 580

Db 90 CysLeuilePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysPro-- 108

QY 579 TGTGTGACTGATGCGACTCTTGCAGATCACCTTCCATGTCGC----- 538

Db 109 LeuileProGlySerSerThrThrSerThrGlyProCysArgAsnThrThrCysThrThr 128

QY 537 ---GCACTGGGATCTCAGCTGTATCTGATCTTCTTCTGCTGCTGCCAGGTGTAGG-- 483

Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148

QY 482 ----ATCCGACTTTCAGCTGCTGCTGCTTC----- 456

Db 148 nCysThrCysileProileProSerSerTrpAlaPheAlaLysPheLeuTrpGluTrpAl 168

QY 455 -----TTTTTGGGGCT 445

Db 168 aSerValArgPheSerTrpLeuSerLeuLeuValProPheValGlnTrpPheValGlyLe 188

QY 444 CTCCATTACTGGCTCTCTGGGACTG-----GCTGTGTGTTCTGTGTGTGTGTGTGTCG 391

Db 188 userProThrValTrpLeuSerValileTrpMetMetTrpTyr----- 202

QY 390 CTCATGCGCTCCCTCTCTGCTGCTGCTGCGGA-----GTCCAGAAATCTTCTGTGCGG 337

Db 203 -----TrpGlyProSerLeuTyraAsnileLeuSerPr 213

QY 336 GTTCCAGGTGCTTCTGGGACTCGGATAC 309

Db 213 oPheMetProLeuLeuProilePheTyr 222

RESULT 19

PCT-US94-05090-2

; Sequence 2, Application PC/TUS9405090

; GENERAL INFORMATION:

; APPLICANT: CARMAN, WILLIAM

; APPLICANT: DECKER, RICHARD H

; APPLICANT: WALLACE, LESLEY

; APPLICANT: MIMMS, LARRY T

; APPLICANT: SOLOMON, LARRY R

; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D

; STREET: ONE ABBOTT PARK ROAD

; CITY: ABBOTT PARK

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/05090
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: POREMSKI, PRISCILLA E.
: REGISTRATION NUMBER: 33,207
: REFERENCE/DOCKET NUMBER: 5347.US.01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 708-937-6365
: TELEFAX: 708-938-2623
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 228 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-05090-2

```

Alignment Scores:		
Pred. No.:	0.348	Length: 228
Score:	88.00	Matches: 45
Percent Similarity:	35.79%	Conservative: 23
Best Local Similarity:	23.68%	Mismatches: 59
Query Match:	6.17%	Indels: 63
DB:	5	Gaps: 9

US-09-658-824-808 (1-781) x PCT-US94-05090-2 (1-228)

QY	744	GATAGTTTAAGTCAAAATATCTATATAAAACACGACGTGGTGTGTTTCACGCTGTG	-----	691
Db	52	AsnSerGlnSerGlnIleSerHisSerProThrCysCysProIleCysProGly	71	
QY	690	-----CTTCATTAAACTTGGTGTGCTCTCCACCTGCTTCCTGGCATTTACAGTG	640	
Db	72	TyArgTrpMetCysLeuArgPheIleIlePheLeuCysIle	-----	LeuLeuLeu 89
QY	639	TTCTCTTTTAGGTATTATCTACCTTGACGGCGGAACCAATCCAGATTATCCCGG	580	
Db	90	CysLeuIlePheLeuValLeuAspTyrGlnGlyMetLeuProValCysPro	-----	108
QY	579	TGTTTGACGATCCAGCTCTTGCGAGATCACCTTCACATGTCG	-----	538
Db	109	LeuIleProGlySerSerThrThrSerThrGlyProCysArgAsnThrThrCysThrThr	128	
QY	537	---GCACTGGGATCTCAGCTGTATCTGATCTCTCTCTGCTGCCAGGTGAGG	---	483
Db	129	ProAlaGlnGlyThrSerMetPheProSerCysCysCysThrLysProThrAspArgAs	148	
QY	482	-----ATCCCGACCTTTCAGCTGCTGGTCTTC	-----	456
Db	148	nCysThrCysIleProIleProSerSerTrpAlaPheAlaLysPheLeuTrpGluTrpAl	168	
QY	455	-----	-----	445
Db	168	aSerValArgPheSerTrpLeuSerLeuLeuValProPheValGlnTrpPheValCylLe	188	
QY	444	CTCCATTACTGGGCTCCTGGGACTG	-----	CCTGTGGTTCGTGTTGTGTGTGTCG 391
Db	188	uSerProThrValTrpLeuSerValIleTrpMetMetTrpTrp	-----	202
QY	390	CTCATCCGCTCCCTCTCTCTGGTCCGCTCGGGGA	-----	GPCCAGAACTTCTCTCTCGCGG 337
Db	203	-----	-----	TrpGlyProSerLeuTrpAsnIleLeuSerPr 213
QY	336	GTTCAGGTGCTTTCGGGACTCGGATAC	309	
Db	213	oPheMetProLeuLeuProIlePheTrp	222	
RESULT 20				

## RESULT 20

```

US-09-463-702A-2
: Sequence 2, Application US/09463702A
: Patent No. 6335435
: GENERAL INFORMATION:
: APPLICANT: AGENE Research Institute, Co., Ltd.
: APPLICANT: HIRAKI AND ASSOCIATES
: APPLICANT: SHIMAMOTO, AKIRO
: APPLICANT: KITAO, SAORI
: APPLICANT: FURUICHI, YASUHIRO
: TITLE OF INVENTION: HUMAN GENE RECO4 ENCODING HELICASE
: FILE REFERENCE: HIRAI150
: CURRENT APPLICATION NUMBER: US/09/463,702A
: CURRENT FILING DATE: 2000-01-24
: PRIOR APPLICATION NUMBER: PCY/JP98/03114
: PRIOR FILING DATE: 1998-07-10
: PRIOR APPLICATION NUMBER: JAPAN 9/200387
: PRIOR FILING DATE: 1997-07-25
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 2
: LENGTH: 1208

```

; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-463-702A-2

Alignment Scores:		
Pred. No.:	0.685	Length: 1208
Score:	88.00	Matches: 58
Percent Similarity:	34.67%	Conservative: 20
Best Local Similarity:	25.78%	Mismatches: 94
Query Match:	6.38%	Indels: 53
DB:	4	Gaps: 10

US-09-658-824-808 (1-781) x US-09-463-702A-2 (1-1208)

QY	66	ACGGCGGTAGTCCACAGCCACATCCAACTGGAGTGTGAAGTGTGAGTGAGTGAGGA	125
DB	47	ThrLeuIysArgThrThrGlyGlnAlaGlyGlyLeu-----Arg	60
QY	126	GGAACACAGCAGGCTCCGGAGGGTTGT--GTGGTCAGTGACTCAGTGAGGAAGGCCCT	182
DB	61	SerSerGluSerLeuProAlaAlaGluGluAlaProGluProArgCysTrpGlyPro	80
QY	183	CGAAGTCGTGTCCTCTCATCGGTGCGCACGCCCATGGACCTCTTGTCTCTGTCACGCG	242
DB	81	HisLeuAsnArgAlaAlaThrLysSerProGlnProThrProGlyArgSerArgGlnGly	100
QY	243	CAT-----AACTAGGAGGAGGCGGAGGAGTGGAGGGCTCAGCCCAAGCTGGG	296
DB	101	SerValProAspTrpGlyGlnArgLeuLysAlaAsnLeuLysGlyThrLeuGlnAlaGly	120
QY	297	GTGCTGTTGGG-----GTATCCGAGTCCTCCAGGAAGCACCTGGAACCC	338
DB	121	ProAlaLeuGlyArgArgProTrpProLeuGlyArgAlaSerSerLysAlaSerThrPro	140
QY	339	CGACACGAAGATTCTGACTCCCCAGACAGGGACCAGGAGGAGCGATGACCGCACAC	398
DB	141	Lys-----ProProGlyThrGlyProValProSerPheAlaGluLys	154
QY	399	ACAAACACAGAACCACACAGCCAGTCCCAGGAGCCCAAGTAAATGGAGAGCCCCCA	458
DB	155	ValSerAspGluPro-----ProGlnLeuPro	163
QY	459	GAACACGACAGTGAAGAGTCGGATCCTACACCTGGGCAG-----CAGACAGAA	506
DB	164	GluPro-----GlnProArgProGlyArgLeuGlnHisLeuGlnAlaSer	178
QY	507	GAAGATCCAGGATACAGCTCAGATCCAGTGCAGACATGGAAAGGTGATCTCGAAGCTG	566
DB	179	LeuSerGlnArgLeuGlySerLeuAsp-----ProGlyTrpLeuGlnArgCys	194
QY	567	CATCAGTCAAAACACCG-----GGGATAAAATCTGGATTTGGGTTC	605

```
Db 195 His-SerGluValProAspPheLeuGlyAlaProLysAlaCysArgProAspLeuGly 214
QY 606 CGCGTCAGGTGAAGATAATACCTAAAGAGGAACACTGTAATAATCCCAAGCAGGTGA 665
Db 214 rGluGluSerGlnLeuIleProGlyGluSerAlaValLeuGlyProGlyAlaGly 234
QY 666 AGAGCAACCAAA 678
Db 234 rGlnGlyProGlu 238
RESULT 21
5164481-2
; Patent No. 5164481
; TITLE OF INVENTION: LACROIX, MARTIAL; ZREIN, MAAN; DIONNE, GERVAIS
; FOR DETECTING AND ELICITING ANTIBODIES TO RUBELLA VIRUS
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/397,767
; FILING DATE: 23-AUG-1989
; SEQ ID NO: 2:
; LENGTH: 277
5164481-2
Alignment Scores:
Pred. No.: 0.426 Length: 277
Score: 87.50 Matches: 40
Percent Similarity: 29.61% Conservative: 5
Best Local Similarity: 26.32% Mismatches: 55
Query Match: 6.14% Indels: 52
DB: 6 Gaps: 7
US-09-658-824-808 (1-781) x 5164481-2 (1-277)
QY 372 TGGTCCCGTCTGGGAGTCCA-----GAATCTTCTGTCGGGTTCCAGGT 328
Db 74 TrpSerArgAlaProProProGluGluArgGlnGluSerArgSerGlnThrProAla 93
QY 327 GCTTCTGGGACTCGATACCCCAACAGCACCC-----CAGCTTCGCCCTG 283
Db 94 ProLysProSerArgAlaProGlnGlnProGlnProProArgMetGlnThrGlyArg 113
QY 282 AGCCCTCCACTCCCTCGGCCCTCTCCCTCCCTAGTTATGCGCGTGACGAGCAAGAAGG 223
Db 114 GlyGlySerAlaProArgProGluLeuGlyProThrAsnProPheGlnAlaVal 133
QY 222 TCATGGGCGTGGCAGCGATGAGAGGAGCAGACTTCGAGGCGCTTCTCAGCTCTGAGT 163
Db 134 AlaArgGlyLeuArgProProLeuHisAspProAspThrGluAla----- 148
QY 162 CACTGACCAACAACCCCTCCGGAAGCCTGC-----TGGTTCTCTC 124
Db 149 -----ProThrGluAlaCysValThrSerTrpLeuTrpSerGluGlyGln 163
QY 123 -----TTCACCTCTCACTCACTTCACTCC----- 97
Db 164 GlyAlaValPheTyrArgValAspLeuHisPheThrAsnLeuGlyThrProProLeuAsp 183
QY 96 -----AGTTGGATCTGCTGGACCTACCCCGCTGTCTCAGTAGCGGAGAAAGA 46
Db 184 GluAspGlyArgTrpAspProAlaLeuMetTyrAsnProCys----- 197
QY 45 ATCCAGACTCAGGAGCCGAGTCGCGGCTCAC 12
Db 198 -----Gly-ProGluProProAlaHis 204
RESULT 22
US-08-127-499A-8
; Sequence 8, Application US/08127499A
; Patent No. 5510264
; GENERAL INFORMATION:
; APPLICANT: VAN ALSTYNE, Diane
; APPLICANT: SHARMA, Lawrence Rajendra
```

```
; TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
; TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/127,499A
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 51916/102/INBI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1063 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; US-08-127-499A-8
Alignment Scores:
Pred. No.: 0.735 Length: 1063
Score: 87.50 Matches: 40
Percent Similarity: 29.61% Conservative: 5
Best Local Similarity: 26.32% Mismatches: 55
Query Match: 6.14% Indels: 52
DB: 1 Gaps: 7
US-09-658-824-808 (1-781) x US-08-127-499A-8 (1-1063)
QY 372 TGGTCCCGTCTGGGAGTCCA-----GAATCTTCTGTCGGGTTCCAGGT 328
Db 74 TrpSerArgAlaProProProGluGluArgGlnGluSerArgSerGlnThrProAla 93
QY 327 GCTTCTGGGACTCGATACCCCAACAGCACCC-----CAGCTTCGCCCTG 283
Db 94 ProLysProSerArgAlaProGlnGlnProGlnProProArgMetGlnThrGlyArg 113
QY 282 AGCCCTCCACTCCCTCGGCCCTCTCTCCCTAGTTATGCGCGTGACGAGCAAGAAGG 223
Db 114 GlyGlySerAlaProArgProGluLeuGlyProThrAsnProPheGlnAlaVal 133
QY 222 TCATGGGCGTGGCAGCGATGAGAGGAGCAGACTTCGAGGCGCTTCTCAGCTCTGAGT 163
Db 134 AlaArgGlyLeuArgProProLeuHisAspProAspThrGluAla----- 148
QY 162 CACTGACCAACAACCCCTCCGGAAGCCTGC-----TGGTTCTCTC 124
Db 149 -----ProThrGluAlaCysValThrSerTrpLeuTrpSerGluGlyGln 163
QY 123 -----TTCACCTCTCACTCACTTCACTCC----- 97
Db 164 GlyAlaValPheTyrArgValAspLeuHisPheThrAsnLeuGlyThrProProLeuAsp 183
QY 96 -----AGTTGGATCTGCTGGACCTACCCCGCTGTCTCAGTAGCGGAGAAAGA 46
Db 184 GluAspGlyArgTrpAspProAlaLeuMetTyrAsnProCys----- 197
QY 45 ATCCAGACTCAGGAGCCGAGTCGCGGCTCAC 12
```

```

Db 198 -----Gly-ProGluProProAlaHis 204
RESULT 23
US-08-482-847-8
; Sequence 8, Application US/08482847
; Patent No. 5556757
; GENERAL INFORMATION:
; APPLICANT: VAN ALSTYNE, Diane
; APPLICANT: SHARMA, Lawrence Rajendra
; TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
; TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
; TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,847
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/127,499
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 51916/104/INBI
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1063 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
US-08-482-847-8
Alignment Scores:
Pred. No.: 0.735 Length: 1063
Score: 87.50 Matches: 40
Percent Similarity: 29.61% Conservative: 5
Best Local Similarity: 26.32% Mismatches: 55
Query Match: 6.14% Indels: 52
DB: 1 Gaps: 7
US-09-658-824-808 (1-781) x US-08-482-847-8 (1-1063)
QY 372 TGGTCCCGTCTGGGAGTCCA-----GAATCTCTCTCGGGGTCCAGGT 328
Db 74 TrpSerArgAlaProProProGluGluArgGlnGluSerArgSerGlnThrProAla 93
QY 327 GCTTCTGGGACTGGGATACCCCAACAGCACCC-----CAGCTTCGCCGTG 283
Db 94 ProLysProSerArgAlaProProGlnGlnProGlnProProArgMetGlnThrGlyArg 113
QY 282 AGCCCTCCACTCCTCGGCCCTCCTCTCCCTAGTTATGGCCGTGACGACAGAAAGG 223
Db 114 GlyLysSerAlaProArgProGluLeuGlyProThrAsnProPheGlnAlaVal 133
QY 222 TCCATGGCGGTGGCACCGCATCAGAGGACGACGACTTCGAGGGCCTTCTCACTCTGAGT 163
::: ::::: ::::: ::::: ::::: :::::

```

```

Db 134 AlaArgGlyLeuArgProProLeuHisAspProAspThrGluAla----- 148
QY 162 CACTGACCACAAACCTCCGGAAGCCTGC-----TGGTTCCTC----- 124
Db 149 -----ProThrGluAlaCysValThrSerTrpLeuTrpSerGluGlyGln 163
QY 123 -----TTCACCTCTCACTCACACTTCACTCCC----- 97
Db 164 GlyAlaValPheTyrArgValAspLeuHisPheThrAsnLeuGlyThrProProLeuAsp 183
QY 96 -----AGTTGGATCTGCCTGTGGACCTACCCGCCGTCTCTCAGTAGCGGAGAAAGA 46
Db 184 GluAspGlyArgTrpAspProAlaLeuMetTyrAsnProCys----- 197
QY 45 ATCCAGACCTCAGGACCCGAGTCGCGCGCTCAC 12
Db 198 -----Gly-ProGluProProAlaHis 204
RESULT 24
US-08-127-499A-1
; Sequence 1, Application US/08127499A
; Patent No. 5510264
; GENERAL INFORMATION:
; APPLICANT: VAN ALSTYNE, Diane
; APPLICANT: SHARMA, Lawrence Rajendra
; TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
; TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/127,499A
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 51916/102/INBI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 992 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
US-08-127-499A-1
Alignment Scores:
Pred. No.: 0.913 Length: 992
Score: 86.50 Matches: 40
Percent Similarity: 29.87% Conservative: 6
Best Local Similarity: 25.97% Mismatches: 52
Query Match: 6.07% Indels: 56
DB: 1 Gaps: 7
US-09-658-824-808 (1-781) x US-08-127-499A-1 (1-992)
QY 372 TGGTCCCGTCTGGGAGTCCA-----GAATCTCTCTCGGGGTCCAGGT 328
Db 73 TrpSerArgAlaProProProGluGluArgGlnGluSerArgSerGlnThrProAla 92

```

```
QY 327 GCTTCTGGACTCGGATACCCCAACAGCACCC-----CAGCTTCGCCTG 283
Db 93 ProLysProSerArgAlaProGlnProGlnProProArgMetGlnThrGlyArg 112
QY 282 AGCCCTCCACTCCTCGGCCCTCCTCCTCCTAGTTATGGCGGTGACGAGACAAGAGG 223
Db 113 GlyGlySerAlaProArgProGluLeuGlyProProThrAsnProPheGlnAlaVal 132
QY 222 TCATGGCGCTGGCAGCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 163
Db 133 AlaArgGlyLeuArgProProLeuHisAspProAspThrGluAla----- 147
QY 162 CACTGACACACACACCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 103
Db 148 -----ProThrGluAlaCys----- 152
QY 102 ACTCCAGTTGGATCTGCCTGTGGACC----- 76
Db 153 ValThrSerTrp-----LeuTrpSerGluGlyAlaValPheTyrArgValasp 170
QY 75 -----TACCGCGCGTGTCTCAGTAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 42
Db 171 LeuHisPheIleAsnLeuGlyThrProProLeuAspGluAspGly-ArgTrpAspProAl 190
QY 41 -----AGACCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12
Db 190 aLeuMetTyrAsnProCysGlyProGluProProAlaHis 203

RESULT 25
US-08-482-847-1
; Sequence 1, Application US/08482847
; Patent No. 5556757
; GENERAL INFORMATION:
; APPLICANT: VAN ALSTYNE, Diane
; APPLICANT: SHARMA, Lawrence Rajendra
; TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
; TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
; TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THEREO, AND USES THEREOF
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,847
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/127,499
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 51916/104/INBI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 992 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
US-08-482-847-1
```

```
Alignment Scores:
Pred. No.: 0.913 Length: 992
Score: 86.50 Matches: 40
Percent Similarity: 29.87% Conservative: 6
Best Local Similarity: 25.97% Mismatches: 52
Query Match: 6.07% Indels: 56
DB: 1 Gaps: 7

US-09-658-824-808 (1-781) x US-08-482-847-1 (1-992)
QY 372 TGTTCCTCGCTGGGAGTCCG-----GAATCTTCTGTCGGGTTCACGAGT 328
Db 73 TrpSerArgAlaProProProGluGluArgGlnGluSerArgSerGlnThrProAla 92
QY 327 GCTTCTGGACTCGGATACCCCAACAGCACCC-----CAGCTTCGCCTG 283
Db 93 ProLysProSerArgAlaProGlnProGlnProProArgMetGlnThrGlyArg 112
QY 282 AGCCCTCCACTCCTCGGCCCTCCTCCTCCTAGTTATGGCGGTGACGAGACAAGAGG 223
Db 113 GlyGlySerAlaProArgProGluLeuGlyProProThrAsnProPheGlnAlaVal 132
QY 222 TCATGGCGCTGGCAGCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 163
Db 133 AlaArgGlyLeuArgProProLeuHisAspProAspThrGluAla----- 147
QY 162 CACTGACACACACACCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 103
Db 148 -----ProThrGluAlaCys----- 152
QY 102 ACTCCAGTTGGATCTGCCTGTGGACC----- 76
Db 153 ValThrSerTrp-----LeuTrpSerGluGlyAlaValPheTyrArgValasp 170
QY 75 -----TACCGCGCGTGTCTCAGTAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 42
Db 171 LeuHisPheIleAsnLeuGlyThrProProLeuAspGluAspGly-ArgTrpAspProAl 190
QY 41 -----AGACCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12
Db 190 aLeuMetTyrAsnProCysGlyProGluProProAlaHis 203

RESULT 26
US-08-882-046-4
; Sequence 4, Application US/08882046
; Patent No. 6136952
; GENERAL INFORMATION:
; APPLICANT: Li, Linheng
; APPLICANT: Hood, Leroy
; APPLICANT: Krantz, Ian D.
; APPLICANT: Spinner, Nancy B.
; TITLE OF INVENTION: Human Jagged Polypeptide, Encoding
; TITLE OF INVENTION: Nucleic Acids and Methods of Use
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,046
; FILING DATE: 25-JUN-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
```

```
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UW 2637
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1148 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-882-046-4

Alignment Scores:
Pred. No.: 0.969 Length: 1148
Score: 86.50 Matches: 37
Percent Similarity: 35.03% Conservative: 18
Best Local Similarity: 23.57% Mismatches: 49
Query Match: 6.07% Indels: 53
DB: 4 Gaps: 10

US-09-658-824-808 (1-781) x US-08-882-046-4 (1-1148)
QY 585 CCCCCTGTTGACGTGACGTCTTCAGATCACCTTCACATGTCGCGACTGGGACT 526
Db 443 ProLeuCysGluValAspValAspLeuCysGluProSerProCysArgAsn---GlyAla 461
QY 525 CAGCTGTAT-----CCTGATCTTCTCTGCTGCTGCC----- 492
Db 462 ArgCysTyrAsnLeuGluGlyAsp-TyrtYrCysAlaCysProAspAspPheGlyGly 481
QY 491 -AGGTGTAGGATCCCGACTTTTCAGCTGCTGCTCTCTCTTT-----TGGGCT 445
Db 481 sasncysSerValProArgGluProCysTrp-ArgGlyLeuInsSerAspArgTrpLeuA 501
QY 444 CTCCTATTACTGGCTCTCGGAC---TCG-----CTGTGCTGTTCTGTGT 403
Db 501 rgValArgArgGlyAlaTrpAspAlaTrpHisSerThrSerGlyValCysGlyProHisG 521
QY 402 TTGTGTGTGCTGCTCATGCC-----GTCCCTCTCTGCTCCGCTCGGGAGTCCA 352
Db 521 LyArgCysValSerGlnProGlyGlyAsnPheSerCysIleCysAspSerGlyPheThrG 541
QY 351 GAATCTTCTCGGGGTTC----- 332
Db 541 lyThrTyrCysHisGluAsnIleAspAspCysLeuGlyGlnProCysArgAsnGlyGlyT 561
QY 331 -----AGGTGCTTC-----TGGGACTCGGATA 310
Db 561 hrCysIleAspGluValAspAlaPheArgCysPheCysProSerClyTrpGluGlyGluL 581
QY 309 CCCCCAACAGCACCCCA-----GCTTCGCTGAGCCCTCCACTCC 269
Db 581 euCysAspThrAsnProAsnAspCysLeuProAspProCysHisser 596

RESULT 27
US-08-072-610-2
; Sequence 2, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,610
; FILING DATE: 19930602
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/07686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pVMB3.3.1
US-08-072-610-2

Alignment Scores:
Pred. No.: 1.18 Length: 1018
Score: 85.50 Matches: 37
Percent Similarity: 40.00% Conservative: 27
Best Local Similarity: 23.12% Mismatches: 59
Query Match: 6.20% Indels: 37
DB: 1 Gaps: 5

US-09-658-824-808 (1-781) x US-08-072-610-2 (1-1018)
QY 311 ATCCGAGTCCAGAGCACCTTGAACCCCGACAGAGATT----- 350
Db 551 ValGlnValProValAlaValGly-ProAlaGlnGluValProThrGluGluLeuMetG 570
QY 351 -CTGACTCTCCAGACGGGACCGAGGAGGCGCATGACGACACACACACA-- 407
Db 570 nLeuGlnGluAspPheGluLeuGlyThrAlaGluAlaProGluGluGlyGluLe 590
QY 408 -----GAACACACACGCCACTCCAGAGGCCAGTATGAGAGCCCCAAAA 454
Db 590 uValLeuGluGlyGluGlyGluProThrGluGluGluProArgGluGlyGluProThrG 610
QY 455 AGAAGACACGACGCTGAAGTGGGATCTTACACCT----- 491
Db 610 uGlyGluValProGluGluGluLeuGluAlaThrProGluAspPheGluLeuGluG 630
QY 492 -----GGGACGACGACAGAGAGAGATCAGGA 517
Db 630 uProThrGlyGluGluValGluGluThrValGluGlyGluGluAlaGluGlyGluG 650
QY 518 TACAGCTGAGATCCCA-----GTGCGCCACATGGAAGTGTATCTGCAAGA 562
Db 650 uValGluValProAlaGluGluValGluGluValGluProAlaGluValGluG 670
QY 563 GCTGCATCAGTCAACACCCGGGGATAATCTGGATTGGTTCGGCTCAAGGTGAAGA 622
Db 670 uValGluGlu-ValProGluGluValGluGluValProAlaGluValGluGluG 690
QY 623 TAATACCTAAAGAGGAACACACTGTAAATGCCAGACGAGGTGAAGACCAACCAAA 678
Db 690 luValProGluGluValGlu-----GluValProGluGluValGluGluValProGlu 707

RESULT 28
```

```
US-08-719-822B-2
; Sequence 2, Application US/08719822B
; Patent No. 5874527
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,822B
; FILING DATE: 09/30/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pVMB3.3.1
US-08-719-822B-2
Alignment Scores:
Pred. No.: 1.18 Length: 1018
Score: 85.50 Matches: 37
Percent Similarity: 40.00% Conservative: 27
Best Local Similarity: 23.12% Mismatches: 57
Query Match: 6.20% Indels: 39
DB: 2 Gaps: 5

US-09-658-824-808 (1-781) x US-08-719-822B-2 (1-1018)
QY 311 ATCCGAGTCCAGAACGACCTGGACCCGACAGAAAGATT----- 350
DB 551 ValGlnValProValAlaValGly-ProAlaGlnGluValProThrGluGluLeu-detG1 570
QY 351 -CTGACTCCCGACGCGGACGAGGACGCGGATGACGACACACAAACACA-- 407
DB 570 nLeuGlnAspAspPheGluLeuGluGlyThrAlaGluAlaProGluGluGlyGluLe 590
QY 408 -----GAACACACACAGCCAGTCCGAGGACCCAGTAAATGAGAGCCCAAAA 454
DB 590 uValLeuGluGlyGluGlyGluProThrGluGluGluProArgGluGlyGluProThrG1 610
QY 455 AGAAGAACACGACGCTGAAGTCGGGATCCCTACACCT----- 491
DB 610 uGlyValProGluGluGluLeuGluAlaThrProGluAspAspPheGluLeuGluG1 630
QY 492 -----GGCAGCAGACAGAGAAGATCAGGA 517

US-09-092-458-2
; Sequence 2, Application US/09092458
; Patent No. 6231861
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,458
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/719,821
; FILING DATE: 09/30/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pVMB3.3.1
US-09-092-458-2
Alignment Scores:
Pred. No.: 1.18 Length: 1018
Score: 85.50 Matches: 37
Percent Similarity: 40.00% Conservative: 27
Best Local Similarity: 23.12% Mismatches: 57
Query Match: 6.20% Indels: 39
DB: 4 Gaps: 5
```







Query Match:		6.09%	Indels:	9
DB:		3	Gaps:	3
US-09-658-824-808 (1-781) x US-08-545-860D-28 (1-1187)				
Qy	320	CCAGAAGCCTGGAAACCCGACAGAGATTCTGGACTCCCGACGGGACCGAGGAGG	379	
Db	502	ProAlaAlaProProGluGlyProArgSerThrGluProProArgArgHisProGluSer	521	
Qy	380	GACGGCATGACGACACACAAACACAGAACACACAGCAGCTCCCGAGGCCAGCTAA	439	
Db	522	LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----	539	
Qy	440	TGGAGAGCCCAAAAA-----GAAGAACACAGCAGCTGAAAGTCGGGATCTCTACAC	490	
Db	540	-----ProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHisPr	557	
Qy	491	TGGGCAGCAGACA---GAAGAAGATCAGGATACAGCTGAGATCCAGTCGCCGACATGA	547	
Db	557	oGlyLysArgSerCysGlnLysSerProAlaGlnGluProProGlnArgGlnThrVa	577	
Qy	548	AGGTGATCTGCAAGAGCTGCATCATCAGTCAACACCCGGGATAAATCTGGATTGGTCCG	607	
Db	577	lGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSe	597	
Qy	608	CGGTCAAGTGAA	620	
Db	597	rLeuGlnGlyGlu	601	
RESULT 34				
PCT-US94-04496-28				
; Sequence 28, Application PC/TUS9404496				
; GENERAL INFORMATION:				
; APPLICANT: Croce, Carlo				
; APPLICANT: Canaani, Eli				
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods				
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias				
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1				
; NUMBER OF SEQUENCES: 86				
; CORRESPONDENCE ADDRESS:				
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &				
; STREET: One Liberty Place, 46th floor				
; CITY: Philadelphia				
; STATE: Pennsylvania				
; COUNTRY: USA				
; ZIP: 19103				
; COMPUTER READABLE FORM:				
; MEDIUM TYPE: Floppy disk				
; COMPUTER: IBM PC compatible				
; OPERATING SYSTEM: PC-DOS/MS-DOS				
; SOFTWARE: PatentIn Release #1.0, Version #1.25				
; CURRENT APPLICATION DATA:				
; FILING DATE: 07-MAR-1996				
; CLASSIFICATION: 435				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: PCT/US94/04496				
; FILING DATE: 22-APR-1994				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: PCT/US92/10930				
; FILING DATE: 09-DEC-1992				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 08/327,392				
; FILING DATE: 19-OCT-1994				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 08/320,559				
; FILING DATE: 11-OCT-1994				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 08/062,443				
; FILING DATE: 14-MAY-1993				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 07/971,094				
; FILING DATE: 30-OCT-1992				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 07/888,839				
; FILING DATE: 27-MAY-1992				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 07/805,093				
; FILING DATE: 11-DEC-1991				
; ATTORNEY/AGENT INFORMATION:				
; NAME: DeLuca Esq., Mark				
; REGISTRATION NUMBER: 33,229				
; REFERENCE/DOCKET NUMBER: TJU-1262				
; TELECOMMUNICATION INFORMATION:				
; TELEPHONE: (215) 568-3100				
; TELEFAX: (215) 568-3439				
; INFORMATION FOR SEQ ID NO: 28:				
; SEQUENCE CHARACTERISTICS:				
; LENGTH: 1187 amino acids				
; TYPE: amino acid				
; TOPOLOGY: linear				
; MOLECULE TYPE: protein				
US-08-545-860D-28				
Alignment Scores:				
Pred: No.:	1.81	Length:	1187	
Score:	84.00	Matches:	30	
Percent Similarity:	42.86%	Conservative:	15	
Best Local Similarity:	28.57%	Mismatches:	51	

Score: 84.00 Matches: 30  
Percent Similarity: 42.86% Conservative: 15  
Best Local Similarity: 28.57% Mismatches: 51  
Query Match: 6.09% Indels: 9  
DB: 5 Gaps: 3

US-09-658-824-808 (1-781) x PCT-US94-04496-28 (1-1187)

QY 320 CCAGAACACCTGGAACCCCGACAGAAAGATTCTGGACTCCCGACAGCGGACGAGAGG 379  
||| |||||  
Db 502 ProAlaAlaProGluGluProArgSerThrGluProArgArgHisProGluSer 521  
||| |||||  
QY 380 GAGGCGATGACGACACACAAACACACAGCCAGTCCAGGACCCAGTAA 439  
||| |||||  
Db 522 LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----- 539  
||| |||||  
QY 440 TGGAGAGCCCAAAAA-----GAAGAACCCAGCAGCTGAAAGTCGGGATCCTACACC 490  
||| |||||  
Db 540 -----ProProLysSerSerSerLysAlaProArgAlaProGluAlaProHisPr 557  
||| |||||  
QY 491 TGGGACGACAGACA---GAAGAACATCAGGATACAGTCCAGTCCAGTCCGACATGGA 547  
||| |||||  
Db 557 oGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnThrVa 577  
||| |||||  
QY 548 AGGTGATCTGCAGAGCTGCATCAGTCAAAACACCCGGGATAAATCTGGATTTCGGTTCCG 607  
||| |||||  
Db 577 lGlyThrLysGlnProLysProValLysAlaSerAlaArgAlaGlySerArgThrSe 597  
||| |||||  
QY 608 GCGTCAAGGTGAA 620  
||| |||||  
Db 597 rLeuGlnGlyGlu 601

RESULT 35

US-08-320-559-26  
; Sequence 26, Application US/08320559  
; Patent No. 5633135  
; GENERAL INFORMATION:  
; APPLICANT: Croce, Carlo  
; APPLICANT: Canaan, Eli  
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for  
; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias  
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the  
; TITLE OF INVENTION: All-1 Region  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/320,559  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/062,443  
; FILING DATE: 14 MAY 1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/971,094  
; FILING DATE: 30-OCT-92  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/888,830  
; FILING DATE: 27-MAY-92  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/805,093  
; FILING DATE: 11-DEC-91  
; ATTORNEY/AGENT INFORMATION:

; NAME: DeLuca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: TJU-0855  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1210 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-320-559-26

Alignment Scores:  
Pred. No.: 1.83 Length: 1210  
Score: 84.00 Matches: 30  
Percent Similarity: 42.86% Conservative: 15  
Best Local Similarity: 28.57% Mismatches: 51  
Query Match: 6.09% Indels: 9  
DB: 5 Gaps: 3

US-09-658-824-808 (1-781) x US-08-320-559-26 (1-1210)

QY 320 CCAGAACACCTGGAACCCCGACAGAAAGATTCTGGACTCCCGACAGCGGACGAGAGG 379  
||| |||||  
Db 525 ProAlaAlaProGluGluProArgSerThrGluProArgArgHisProGluSer 544  
||| |||||  
QY 380 GAGGCGATGACGACACACAAACACACAGCCAGTCCAGGACCCAGTAA 439  
||| |||||  
Db 545 LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----- 562  
||| |||||  
QY 440 TGGAGAGCCCAAAAA-----GAAGAACCCAGCAGCTGAAAGTCGGGATCCTACACC 490  
||| |||||  
Db 563 -----ProProLysSerSerSerLysAlaProArgAlaProGluAlaProHisPr 580  
||| |||||  
QY 491 TGGGACGACAGACA---GAAGAACATCAGGATACAGTCCAGTCCAGTCCGACATGGA 547  
||| |||||  
Db 580 oGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnThrVa 600  
||| |||||  
QY 548 AGGTGATCTGCAGAGCTGCATCAGTCAAAACACCCGGGATAAATCTGGATTTCGGTTCCG 607  
||| |||||  
Db 600 lGlyThrLysGlnProLysProValLysAlaSerAlaArgAlaGlySerArgThrSe 620  
||| |||||  
QY 608 GCGTCAAGGTGAA 620  
||| |||||  
Db 620 rLeuGlnGlyGlu 624

RESULT 36

US-08-545-860D-26  
; Sequence 26, Application US/08545860D  
; Patent No. 6040140  
; GENERAL INFORMATION:  
; APPLICANT: Croce, Carlo  
; APPLICANT: Canaan, Eli  
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region  
; NUMBER OF SEQUENCES: 94  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &  
; ADDRESSEE: No. 6040140ris  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/545,860D  
FILING DATE: 07-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496  
FILING DATE: 22-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/10930  
FILING DATE: 09-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/327,392  
FILING DATE: 19-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/320,559  
FILING DATE: 11-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/062,443  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,094  
FILING DATE: 30-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/888,839  
FILING DATE: 27-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,093  
FILING DATE: 11-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca Esq., Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-1262  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-860D-26

Alignment Scores:  
Pred. No.: 1.83 Length: 1210  
Score: 84.00 Matches: 30  
Percent Similarity: 42.86% Conservative: 15  
Best Local Similarity: 28.57% Mismatches: 51  
Query Match: 6.09% Indels: 9  
DB: 3 Gaps: 3

US-09-658-824-808 (1-781) x US-08-545-860D-26 (1-1210)

QY 320 CCAGAGACCTGGAACCCGACAGAGATTCTGGACTCCCGACGGGACGAGAGG 379  
DB 525 ProAlaAlaProGluGlyProArgSerThrGluProProArgHisProGluSer 544  
QY 380 GAGCGCATGACGACACACACAGAACACAGACCCAGTCCCGACGGGACGAGTAA 439  
DB 545 LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----- 562  
QY 440 TGGAGAGCCCAAAAA-----CAAGNACCGACGCTGAAGTCGGGATCCTACACC 490  
DB 563 -----ProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHisPr 580  
QY 491 TGGGACGACAGACA---GAAGAAGATCAGGATACAGTACGATCCCGCGGACATGGA 547  
DB 580 oGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnThrVa 600  
QY 548 AGGTGATCTCGAAGACTGCATCAGTCAAAACCGGGGATAATCTGGATTGGGTTCGG 607  
DB 600 lGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSe 620  
QY 608 GCCTCAAGGTGAA 620

Db 620 rLeuGlnGlyGlu 624

RESULT 37

PCT-US94-04496-26  
Sequence 26, Application PC/TUS9404496  
GENERAL INFORMATION:  
APPLICANT: Croce, Carlo  
APPLICANT: Canaan, Eli  
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &  
ADDRESSEE: Norris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496  
FILING DATE:

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca Esq., Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-1242  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-04496-26

Alignment Scores:  
Pred. No.: 1.83 Length: 1210  
Score: 84.00 Matches: 30  
Percent Similarity: 42.86% Conservative: 15  
Best Local Similarity: 28.57% Mismatches: 51  
Query Match: 6.09% Indels: 9  
DB: 3 Gaps: 3

US-09-658-824-808 (1-781) x PCT-US94-04496-26 (1-1210)

QY 320 CCAGAGACCTGGAACCCGACAGAGATTCTGGACTCCCGACGGGACGAGAGG 379  
DB 525 ProAlaAlaProGluGlyProArgSerThrGluProProArgHisProGluSer 544  
QY 380 GAGCGCATGACGACACACACAGAACACAGACCCAGTCCCGACGGGACGAGTAA 439  
DB 545 LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----- 562  
QY 440 TGGAGAGCCCAAAAA-----GAAGNACCGACGCTGAAGTCGGGATCCTACACC 490  
DB 563 -----ProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHisPr 580  
QY 491 TGGGACGACAGACA---GAAGAAGATCAGGATACAGTACGATCCCGCGGACATGGA 547  
DB 580 oGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnThrVa 600  
QY 548 AGGTGATCTCGAAGACTGCATCAGTCAAAACCGGGGATAATCTGGATTGGGTTCGG 607  
DB 548 AGGTGATCTCGAAGACTGCATCAGTCAAAACCGGGGATAATCTGGATTGGGTTCGG 607

```
Db 600 lGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSe 620
QY 608 GCCTCAAGTGAA 620
Db 620 rLeuGlnGlyGlu 624
RESULT 38
US-08-447-591-3
; Sequence 3, Application US/08447591
; Patent No. 5591440
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLACE, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,591
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/059,031
; FILING DATE: 07-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5347.US.01
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-447-591-3
Alignment Scores:
Pred. No.: 1.18 Length: 228
Score: 83.00 Matches: 44
Percent Similarity: 35.26% Conservative: 23
Best Local Similarity: 23.16% Mismatches: 60
Query Match: 5.82% Indels: 63
DB: 1 Gaps: 9
US-09-658-824-808 (1-781) x US-08-447-591-3 (1-228)
QY 744 GATAGTTTAAGTCAATATCTAATATAAACGAGCTGGCTTGTTCACCTTGT----- 591
Db 52 AsnSerGlnThrGlnIleSerHisSerProThrCysCysProProIleCysProGly 71
QY 690 -----CTTCATTAAACTGTGGTTCCTTCCACCTGCTTCTGGCATTTTACAGTG 640
Db 72 TyrArgTipMetCysLeuArgArgPheIleIlePheLeuCysIle-----LeuLeuLeu 89
QY 639 TTCCTCTTTAGGTATTATCTTCACCTTGACCGCGGAAACCAATCCAGATTATCCCGG 580
Db 90 CysLeuIlePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysPro--- 108
QY 579 TGTTTGATGATGACAGCTCTTGCAGATACACCTTCATGTGCG----- 538
Db 109 LeuIleProGlySerSerThrThrSerThrGlyProCysArgAsnThrThrCysThrThr 128
QY 537 ---GCACTGGGATCTCAGCTGTATCTGATCTTCTCTCTGCTGCTGCCAGGTAGG-- 483
Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148
QY 482 -----ATCCCGACCTTCAGCTGCTGGTCTTC----- 45
Db 148 nCysThrCysIleProIleProSerSerTrpAlaPheValLysPheLeuTrpGluTrpAl 168
QY 455 -----TTTTGGGGCT 445
Db 168 aSerValArgPheSerTrpLeuSerPheLeuValProIleValGlnTrpPheAlaGlyLe 188
QY 444 CTCATTACTGGCTCTCTGGGACTG-----GCTGTGTGTTCTGTCTTTGTGTGTCTG 391
Db 188 uSerProThrValTrpLeuSerValIleTrpMetMetTrpTrpYr----- 202
QY 390 CTCATGCCGTCCCTCTCTCTGGTCCCTCTGGGA-----GTCCAGAATCTTCTGTGCGG 337
Db 203 -----TrpGlyProSerLeuTrpAsnIleLeuSerPr 213
QY 336 GTTCCAGGTGCTTCTGGGACTCGGATAC 309
Db 213 oPheMetProLeuLeuProIlePheTrp 222
RESULT 39
US-08-450-943-3
; Sequence 3, Application US/08450943
; Patent No. 5593825
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLIS, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR
; TITLE OF INVENTION: DETECTION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,943
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5347.US.01
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```



Search completed: October 18, 2002, 09:53:09  
Job time : 33.5 secs



Result No.	Score	Query %			DB	ID	Description
		Match	Length	Length			
1	113	8.2	908	2	A33280	sarcalumenin precu	
2	106	7.7	230	2	A29573	Glx-rich protein -	
3	106	7.7	246	2	A38647	glutamine/glutam	
4	106	7.7	255	2	B36477	glutamine/glutam	
5	106	7.7	527	2	A32469	80K protein H prec	
6	104	7.5	319	2	F75420	hypothetical prote	
7	103	7.5	247	2	A29545	submandibular glan	
8	103	7.2	363	1	CBUTB	ubiquinol--cytochr	
9	101	7.3	1546	1	CGHU2E	collagen alpha 2(X	
10	99.5	7.0	598	2	A37251	probable nuclear h	
11	98	7.1	299	2	A57652	cyclic nucleotide-	
12	97	6.8	371	1	H22848	ubiquinol--cytochr	
13	96.5	7.0	277	2	S78063	homeobox protein p	
14	96.5	7.0	815	2	B30843	glutinin high mole	



Qy 178 GCCCTCGAAGTCGTCCTCTCATCGCTGCCAGCCCATGGACCTCTCTGTCCTGTC 237  
Db ||| |||||:||||| ||| :|||  
Qy 345 ProProGluGluGlnValProValMetArgGlnAspGluGlyGluAla----- 361  
Qy 238 ACGGCCATAACTAGGAGGAGGAGGCGCGAGAGTGAGGGCTCAGCGAAGCTGGGG 297  
Db :||| ||||| ||| |||||:||||| :|||  
Qy 362 -----SerSerGluGluGluGlyAspGluGlyGlySerGluGluGlu----- 376  
Qy 298 TGCTGTTGGGGTATCCGAGTCCAGAA-----GCACCTCGAACCCTCC 339  
Db ||| ||||| ||| |||  
Qy 377 -----GluGlyAspProSer-GluGluAspSerGlyGluAspSerGlyAlaLase 394  
Qy 340 GACAGAAGATTCTGGACTCCCGACAGCGGACAGGAGGAGCGGCGCATGAGCGACACA 399  
Db :|||:||||| :||| :||| |||||  
Qy 394 rSerGluGluAlaGlyAlaAlaSerGluGluAlaSer-GlyThrAlaGlyLeuGlyGluG 414  
Qy 400 CAACACAGAACCCAC-----AGCCACTCCCGAG 429  
Db |||:||||| ||| |||||:|||||  
Qy 414 luGluThrGlnProSerThrGluGlyLeuAspSerGlyProAlaGlySerGlnAlaGlnA 434  
Qy 430 AGCCCAAGTAATGGAGAGCCCAAGAAAGAACAGCAGCGTGAAGTCCGGATCTACAC 489  
Db :||| |||||:||||| |||||  
Qy 434 spThrGluAlaGluProGluGluGlyHisGlnGlyProGluSerProIleThrAlap 454  
Qy 490 GTGGCCACACAGAAAGATCAGGATACAGCTGAGATCCAGTCGCGGACATGGAAG 549  
Db || :|||:||||| :||| |||||:|||||  
Qy 454 roGlnGluGluThrGluAspValSerGlu-----GluValProMetArgasp----- 469  
Qy 550 GTGATCTCAAGAGCTGCATGCTCAACACACCGGGGATATCTGATTTGGTTCGCGC 609  
Db ----- 469  
Qy 610 GTCAGGTGAGATATACCTAAGAGGACACTGTAATGTCAGAGCGAGGTGAAGAG 669  
Db -----ArgSerHisIleGluL 475

## RESULT 2

A29573  
Glx-rich protein - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 11-Jan-2000  
C:Accession: A29573  
R:Mirreis, L.; Bedl, G.S.; Dickinson, D.P.; Gross, K.W.; Tabak, L.A.  
J. Biol. Chem. 262, 7289-7297, 1987  
A:Title: Molecular characterization of glutamic acid/glutamine-rich secretory proteins  
A:Reference number: A29573; MUID:87222334  
A:Accession: A29573  
A:Molecule type: mRNA  
A:Residues: 1-230 <MIR>  
A:Cross-references: GB:J02730; NID:g204476; PIDN:AAA1276.1; PID:g204477  
A:Experimental source: submandibular gland  
A:Superfamily: proline-rich protein  
C:Keywords: saliva; submandibular gland

Alignment Scores:  
Pred. No.: 0.344 Length: 230  
Score: 106.00 Matches: 40  
Percent Similarity: 40.29% Conservative: 16  
Best Local Similarity: 28.78% Mismatches: 50  
Query Match: 7.68% Indels: 33  
DB: 2 Gaps: 6

US-09-658-824-808 (1-781) x A29573 (1-230)

Qy 306 GGGGTATCCGATCCAGAGACCTGGAAACCCCGACAGAAAGATTCTGGACTCCCGACAG 365  
Db ||| |||:||||| ||||| |||  
Db 58 GlySerGluGluGlnGlnGlnGlnGlnGluProThrGlnAlaGluAsnGlnGluPro 77

Qy 366 GGGACAGAGAGGAGCGGATGACGACACACACAACACAGAACACAGCCAGTCC 425  
Db ||| ||| :||| |||  
Qy 78 AlaThr-----SerGlySerGluGluGluGlnGln 88  
Qy 426 CAGGAGCCAGN-----AATGGAGAGCCC----- 449  
Db |||||:||||| ||| |||||  
Qy 89 GlnGluProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySerGluGlu 108  
Qy 450 ---CAAAAAGAAACACAGCA---GCTGAAAGTCGGGATCTTACACCTGGCGGACGAGACA 503  
Db |||||:||||| |||||:||||| :|||  
Qy 109 GlnGlnGlnGluProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySer 128  
Qy 504 GAAGAAGATCAGATACACTGAGATCCAGTCGCGACATGGAAGGTGATCTGCAAGAG 563  
Db |||||:||||| ||| ||| :||| |||||  
Qy 129 GluGluGluGlnGlnGlnGlu-----SerThrGlnAlaGluAsnGlnGlu 144  
Qy 564 CTGCATCAGTCAACACCGCGGATATAATCTGGATTGGTTCGCGCTCAAGGTGAAGAT 623  
Db ||| :||| ||| ||| :|||  
Qy 145 -----ProSerAspSerAlaGlyGluGlyGlnGluThrGlnProGluGlu 159  
Qy 624 AATACCTAAGAGGAACACTGTAAATGCCAGNAGCAGGTGAAGACCAACCAA 678  
Db ||| ||||| :||| |||||  
Qy 160 GlyAsn-ValGluSerProProSerSerProGluAsnSerGlnGluGlnProGln 177  
RESULT 3  
A38647  
glutamine/glutamate-rich protein form A, submandibular gland - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 21-Feb-1992 #sequence\_revision 10-Apr-1992 #text\_change 24-Nov-1999  
C:Accession: A38647  
R:Cooper, L.F.; Elia, D.M.; Tabak, L.A.  
J. Biol. Chem. 266, 3532-3539, 1991  
A:Title: Secretagogue-coupled changes in the expression of glutamine/glutamic acid-ri  
A:Reference number: A38647; MUID:91139639  
A:Accession: A38647  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-246 <COO>  
A:Cross-references: GB:M58653; NID:g204481; PIDN:AAA1278.1; PID:g204482  
A:Note: the authors translated the codon GAT for residue 135 as Asn, and CGA for resi  
C:Superfamily: proline-rich protein

## Alignment Scores:

Pred. No.: 0.346 Length: 246  
Score: 106.00 Matches: 41  
Percent Similarity: 37.88% Conservative: 9  
Best Local Similarity: 31.06% Mismatches: 63  
Query Match: 7.68% Indels: 19  
DB: 2 Gaps: 4

US-09-658-824-808 (1-781) x A38647 (1-246)

Qy 306 GGGTATCCGATCCAGAGACCTGGAAACCCCGACAGAAAGATTCTGGACTCCCGACAG 365  
Db ||| ||| :||| |||  
Db 74 GlySerGluGluGlnGlnGlnGlnGluProThrGlnAlaGluAsnGlnGluProPro 93  
Qy 366 GGGACAGAGAGGAGCGGATGACGACACACACAACACAGAACACACAGCCAGTCC 425  
Db ||| ||| :||| |||  
Qy 94 AlaThr-----SerGlySerGluGluGlnGlnGln 104  
Qy 426 CAGGAGCCAGTATATGAGAGAGCCCAAGAAAGAACAGACAGCTGAAGTCCGGATCCT 485  
Db |||||:||||| ||| ||| ||||| |||  
Qy 105 GlnGluProThrGlnAlaGluAsnGln-----GluProProAlaThrSerGlySerGlu 122  
Qy 486 ACACCTGGGACGACAGACAGAA-----GAAGATCAGGTATACAGTGTGATCCCA 533  
Db ||||| ||| ||||| |||||  
Qy 123 GluGluGlnGlnGlnGlnGluProThrGlnAlaGluAspGlnGlnProProAlaThrSer 142  
Qy 534 GTGCCGCACATGGAGGTGATCTGCAAGAGCTGCATCAGTCAACACACC-----GGG 584  
Db :||| ||| ||||| |||||  
Qy 143 GlySerGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 162  
Qy 585 GATAAATCTGGATTGGGTTCCGCGCTCAAGGTGAAGATAATACCTAAAGAGGAACACTG 644

Db 163 AspSerAlaGluGluGlnGlnGluThrGlnProGluGluGlyAsn-VaIGluSerProPr 182  
QY 645 TAAATGCCAGAACGAGGTGAAGACGACCAACCAAC 678  
Db 182 oSerSerProGluAsnSerGlnGluGlnProGln 193

RESULT 4  
B38647  
glutamine/glutamate-rich protein form B, submandibular gland - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Feb-1992 #sequence\_revision 10-Apr-1992 #text\_change 11-Jan-2000  
C:Accession: B38647

R:Cooper, L.F.; Elita, D.M.; Tabak, L.A.  
J. Biol. Chem. 266, 3532-3539, 1991  
A:Title: Secretagogue-coupled changes in the expression of glutamine/glutamic acid-rich  
A:Reference number: A38647; MUID:91139639  
A:Accession: B38647  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-255 <COO>  
A:Cross-references: GB:M58654; NID:G204483; PIDN:AAA1279.1; PID:G204484  
A:Note: the authors translated the codon GAT for residue 135 as Asn, and GAT for residue  
C:Superfamily: proline-rich protein

Alignment Scores:  
Pred. No.: 0.347 Length: 255  
Score: 106.00 Matches: 41  
Percent Similarity: 37.88% Conservative: 9  
Best Local Similarity: 31.06% Mismatches: 63  
Query Match: 7.68% Indels: 19  
DB: 2 Gaps: 4

US-09-658-824-808 (1-781) x B38647 (1-255)

QY 306 GGGGTATCCGAGTCCAGAACACCTGGAAACCCGACAGAGATTCTGGACTCCCGAC 365  
Db 74 GlySerGluGluGlnGlnGlnGlnGlnGluProThrGlnAlaGluAsnGlnGluPro 93  
QY 366 GGGACAGGAGAGGCGGCATCAGCGACACACAAACACAGAACACACACAGCCAGTCC 425  
Db 94 AlaThr-----SerGlySerGluGluGlnGlnGln 104  
QY 426 CAGGAGCCAGTAATGAGAGCCCAAGAACAGACACAGCTGAAGTCCGGATCCT 485  
Db 105 GlnGluProThrGlnAlaGluAsnGln-----GluProAlaThrSerGlySerGlu 122  
QY 486 ACACCTGGCGAGCAGACAGAA-----GAAGATCAGGATACAGCTGAGATCCCA 533  
Db 123 GluGluGlnGlnGlnGlnGluProThrGlnAlaGluAspGlnGlnProAlaThrSer 142  
QY 534 GTGCGGACATGGAAGTGATCTGCAAGAGCTGCATCAGTCAACACCC-----GGG 584  
Db 143 GlySerGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 162  
QY 585 GATAAATCTGGATTGGTTCCGGCGCTCAAGGTGAAGATAATACCTAAAGAGCAACACTG 644  
Db 163 AspSerAlaGlyGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 182  
QY 645 TAAATGCCAGAACGAGGTGAAGACCAACCAAC 678  
Db 182 oSerSerProGluAsnSerGlnGlnGlnProGln 193

RESULT 5  
A32469  
80K protein H precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 21-Jul-2000  
C:Accession: A32469  
R:Sakai, K.; Hirai, M.; Minoshima, S.; Kudoh, J.; Fukuyama, R.; Shimizu, N.  
Genomics 5, 309-315, 1989  
A:Title: Isolation of cDNAs encoding a substrate for protein kinase C: nucleotide sequen  
A:Reference number: A32469; MUID:90007553

A:Accession: A32469  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-527 <SAK>  
A:Cross-references: GB:J03075; NID:g182854; PIDN:AAA52493.1; PID:g182855  
C:Genetics:  
A:Gene: GDB:PRKCSH; G19P1  
A:Cross-references: GDB:119961; OMIM:177060  
A:Map position: 19p13.2-19p13.2  
C:Keywords: phosphoprotein  
F:1-14/Domain: signal sequence #status predicted <SIG>  
F:524-527/Region: endoplasmic reticulum retention signal

Alignment Scores:  
Pred. No.: 0.365 Length: 527  
Score: 106.00 Matches: 52  
Percent Similarity: 37.74% Conservative: 28  
Best Local Similarity: 24.53% Mismatches: 76  
Query Match: 7.68% Indels: 56  
DB: 2 Gaps: 8

US-09-658-824-808 (1-781) x A32469 (1-527)

QY 52 TCCGCTACTCAGACACGGGGTAGGTCCACAGCAGATCCAACTGGGAGTTGAGTGTG 111  
Db 230 SerValThrGluLeu---GlnThrHisProGluLeuAspThrAspGlyAspGlyAlaLeu 248  
QY 112 AGTGAGAGTGAA-----GAGGAACACGACGAGGCTTC 141  
Db 249 SerGluAlaGluAlaGlnAlaLeuSerGlyAspThrGlnThrAspAlaThrSerPhe 268  
QY 142 CGGAGGTTGTGTGTCAGTGCAGTGCAGAGTGCAGAGGCCCTCGAAGTCGTCGCCCTCTC 201  
Db 269 TyrAspArgValTrpAlaAla-IleArgAspLysTyrArgSerGlu-----283  
QY 202 ATCGGTGCCACGCCCATGGACCTTCTGTCTCGTCACGGCCATACTAGGAGG---AA 258  
Db 284 -----AlaLeuProThrAspLeuProAlaProSerAlaProAspLeuThrGluProLy 301  
QY 259 GGAGGCCGAGGAGTGGAGGGGCTCAGCGGAGCTGGGTGCTGTTGGGGGTATCCGAGT 318  
Db 301 sGluGluGln-----304  
QY 319 CCAGAGACACCTGGAAACCCGACAGAGATTTCTGGACTCCCGACGCGGACAGGAGAG 378  
Db 305 -ProProValProSerSerProThrGluGluGlu-----GluGluGluGluGlu 321  
QY 379 GGACGGCATGAGCAGACACACAAACACACAGACCCAGTCCCGAGGCCAGTA 438  
Db 321 uGluGluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 341  
QY 439 ATGGAGGCCCAAGAAAGAACACAGCAGCTGAAAGTCGGGATCCTACACCTGGGAGC 498  
Db 341 euSerProGlnProAlaSerProAlaGluGluAspLysMetProTyroAspGlu 361  
QY 499 AGCAGAGAGAGATCAGGATACA-----GCTCAGA 528  
Db 361 InThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGlu 381  
QY 529 TCCGAGTGCAGCATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATA 588  
Db 381 rgSerLeuLysAspMetGluGluSerIleArgAsnLeuGluGln-----G 396  
QY 589 AATCTGGATTTGGTTCCGGCGCTCAAGGTGAA 620  
Db 396 luIleSerPheAspPheGlyProAsnGlyGlu 406

RESULT 6  
F75420  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: F75420

A:Map position: 1

Qy	306	GGGTATCCGAGTCCCAAGAACCCGAGACAGAGANTCTGCAGTCCCCAGAC	365
Db	74	GlySerGluGluGluGlnGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGlnPro	93
Qy	366	GGGACCAGGAGAGGACGGCATGATGGACACACACAAACACAGACCCACAGCCAGTCC	425
Db	94	AlaThr-----SerGlySerGluGluGlnGlnGln	104

QY 426 CAGGAGCCCCACT-----AATGAGAGAGCCC----- 449

Db	105	GlnGluProThrGlnAlaGluAsnGlnGluProAlaThrSerGlySerGluGlu	124
Qy	450	--CAAAAGAACACGCA--GCTGAAGTCGGATCTACACTGGCAGCAGACA	503
Db	125	GlnGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGluProAlaThrSerGlySer	144

504 GAAGAGATCAGGATACAGCTGAGATCCCACTGCCGACATGGAAAGCTGATCTGCAAGAG 563

Db	145	GluGluGluGlnGlnGlnGlu-----SerThrGlnAlaGluAsnGlnGlu	160
QY	564	CTGCATCAGTCAAAACACCGGGGATAAATCTGGATTCTGGCGCTCAAGGTCAGAT	623

Db 161 -----ProSerAspSerAlaGlyGluGlyGlnGluThrGlnProGluGlu 175

QY 624 AATACCTTAAGAGGAAACACTGTAAATGCCAGAGCAGTGAAGAGCAACCAAA 678  
 176 GLVAsn-ValGluSerProSerSerProGluAAsnSerGlnGluGlnProGln 193

## RESULT 8

ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome b - Trypanosoma  
C; Species: mitochondrion Trypanosoma brucei

C;Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 03-Mar-2000  
C;Accession: A00160

R;Benne, R.; De Vries, B.F.; Van den Burg, J.; Klaver, B. Nucleic Acids Res. 11, 6925-6941, 1983

A;Reference number: A00160; MOLD:8404.1834  
A;Accession: A00160

A;Molecule type: DNA  
A;Residues: 1-363 <BEN>  
A;Cross-references: GB:X00017; NID:g13737; PIDN:CAA24915.1; PID:g578828  
A;Note: the authors translated the initiation codon ATT for residue 1 as I

C:Comment: The cytochrome b gene was isolated from a 20-kb maxicircle.  
C:Genetics:

A;Genome: mitochondrion  
A;Genetic code: SGC6  
A;Start codon: ATT

**C;Function:** A;Description: the net reaction catalyzed by the ubiquinol--cytochrome-c reductase complex is the reduction of cytochrome c (oxidized) to cytochrome c (reduced) with two hydrogen ions taken up from the mitochondrial matrix and four hydrogens released into the intermembrane space. A;Pathway: oxidative phosphorylation; respiratory chain

**C;Keywords:** chromoprotein; electron transfer; heme; iron; membrane protein

F; 2-334/Domain: cytochrome b homology <CBn>  
F; 2-202/Domain: cytochrome b6 homology <CB6>  
F; 216-334/Domain: plastoquinol--plastocyanin reductase 17K protein homology



QY 453 AAAGAGAACGACGAGCTGAAGTCGGGATCCTACAGCTGGCGACGACAGACAGAGAGAT 512  
Db 202 AsnGlnGluPro-----GlnSerGlnAspProThrProGlyGluGlu-----GluGluIle 218  
QY 513 CAGGATACAGCTGAGATCCCA 533  
Db 219 LeuGluSerSerLeuLeuPro 225  
RESULT 10  
A37251  
N:Alternate names: probable nuclear hormone receptor NAK1 - human  
C:Species: Homo sapiens (man)  
C:Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 20-Jun-2000  
R:Accession: A37251; A61191; A61451  
R:Nakai, A.; Kartha, S.; Sakurai, A.; Toback, F.G.; DeGroot, L.J.  
Mol. Endocrinol. 4, 1438-1443, 1990  
A:Title: A human early response gene homologous to murine nur77 and rat NGFI-B, and related to the chicken c-fos gene  
A:Reference number: A37251; MUID: 91133413  
A:Accession: A37251  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-598 <NAK>  
A:Cross-references: GB:D49728; NID:g1813881; PIDN:BAA08565.1; PID:g1813882  
A:Note: it is uncertain whether Met-1 or Met-35 is the initiator  
R:Bondy, G.P.  
Cell Growth Differ. 2, 203-208, 1991  
A:Title: Phorbol ester, forskolin, and serum induction of a human colon nuclear hormone receptor  
A:Reference number: A61191; MUID: 91329278  
A:Accession: A61191  
A:Molecule type: mRNA  
A:Residues: 35-252, 'I', 254-598 <BON>  
R:Chang, C.; Kokontis, J.; Liao, S.; Chang, Y.  
J. Steroid Biochem. 34, 391-395, 1989  
A:Title: Isolation and characterization of human TR3 receptor: a member of steroid receptor superfamily  
A:Reference number: A61451; MUID: 90173205  
A:Accession: A61451  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-261, 'P', 263-369, 'L', 371-598 <CHA>  
A:Cross-references: GB:L13740; NID:g292833; PIDN:AAA36763.1; PID:g292834  
C:Genetics:  
A:Gene: GDB:HRM  
A:Cross-references: GDB:I25357; OMIM:139139  
A:Map position: 12q13-12q13  
C:Superfamily: probable hormone receptor N10, nuclear; erba transforming protein homologous to src, v-src, v-fes, v-fms, v-abl, v-lmo, v-yes, v-csk, v-tyk, v-tyk2, v-tyk3, v-tyk4, v-tyk5, v-tyk6, v-tyk7, v-tyk8, v-tyk9, v-tyk10, v-tyk11, v-tyk12, v-tyk13, v-tyk14, v-tyk15, v-tyk16, v-tyk17, v-tyk18, v-tyk19, v-tyk20, v-tyk21, v-tyk22, v-tyk23, v-tyk24, v-tyk25, v-tyk26, v-tyk27, v-tyk28, v-tyk29, v-tyk30, v-tyk31, v-tyk32, v-tyk33, v-tyk34, v-tyk35, v-tyk36, v-tyk37, v-tyk38, v-tyk39, v-tyk40, v-tyk41, v-tyk42, v-tyk43, v-tyk44, v-tyk45, v-tyk46, v-tyk47, v-tyk48, v-tyk49, v-tyk50, v-tyk51, v-tyk52, v-tyk53, v-tyk54, v-tyk55, v-tyk56, v-tyk57, v-tyk58, v-tyk59, v-tyk60, v-tyk61, v-tyk62, v-tyk63, v-tyk64, v-tyk65, v-tyk66, v-tyk67, v-tyk68, v-tyk69, v-tyk70, v-tyk71, v-tyk72, v-tyk73, v-tyk74, v-tyk75, v-tyk76, v-tyk77, v-tyk78, v-tyk79, v-tyk80, v-tyk81, v-tyk82, v-tyk83, v-tyk84, v-tyk85, v-tyk86, v-tyk87, v-tyk88, v-tyk89, v-tyk90, v-tyk91, v-tyk92, v-tyk93, v-tyk94, v-tyk95, v-tyk96, v-tyk97, v-tyk98, v-tyk99, v-tyk100, v-tyk101, v-tyk102, v-tyk103, v-tyk104, v-tyk105, v-tyk106, v-tyk107, v-tyk108, v-tyk109, v-tyk110, v-tyk111, v-tyk112, v-tyk113, v-tyk114, v-tyk115, v-tyk116, v-tyk117, v-tyk118, v-tyk119, v-tyk120, v-tyk121, v-tyk122, v-tyk123, v-tyk124, v-tyk125, v-tyk126, v-tyk127, v-tyk128, v-tyk129, v-tyk130, v-tyk131, v-tyk132, v-tyk133, v-tyk134, v-tyk135, v-tyk136, v-tyk137, v-tyk138, v-tyk139, v-tyk140, v-tyk141, v-tyk142, v-tyk143, v-tyk144, v-tyk145, v-tyk146, v-tyk147, v-tyk148, v-tyk149, v-tyk150, v-tyk151, v-tyk152, v-tyk153, v-tyk154, v-tyk155, v-tyk156, v-tyk157, v-tyk158, v-tyk159, v-tyk160, v-tyk161, v-tyk162, v-tyk163, v-tyk164, v-tyk165, v-tyk166, v-tyk167, v-tyk168, v-tyk169, v-tyk170, v-tyk171, v-tyk172, v-tyk173, v-tyk174, v-tyk175, v-tyk176, v-tyk177, v-tyk178, v-tyk179, v-tyk180, v-tyk181, v-tyk182, v-tyk183, v-tyk184, v-tyk185, v-tyk186, v-tyk187, v-tyk188, v-tyk189, v-tyk190, v-tyk191, v-tyk192, v-tyk193, v-tyk194, v-tyk195, v-tyk196, v-tyk197, v-tyk198, v-tyk199, v-tyk200, v-tyk201, v-tyk202, v-tyk203, v-tyk204, v-tyk205, v-tyk206, v-tyk207, v-tyk208, v-tyk209, v-tyk210, v-tyk211, v-tyk212, v-tyk213, v-tyk214, v-tyk215, v-tyk216, v-tyk217, v-tyk218, v-tyk219, v-tyk220, v-tyk221, v-tyk222, v-tyk223, v-tyk224, v-tyk225, v-tyk226, v-tyk227, v-tyk228, v-tyk229, v-tyk230, v-tyk231, v-tyk232, v-tyk233, v-tyk234, v-tyk235, v-tyk236, v-tyk237, v-tyk238, v-tyk239, v-tyk240, v-tyk241, v-tyk242, v-tyk243, v-tyk244, v-tyk245, v-tyk246, v-tyk247, v-tyk248, v-tyk249, v-tyk250, v-tyk251, v-tyk252, v-tyk253, v-tyk254, v-tyk255, v-tyk256, v-tyk257, v-tyk258, v-tyk259, v-tyk260, v-tyk261, v-tyk262, v-tyk263, v-tyk264, v-tyk265, v-tyk266, v-tyk267, v-tyk268, v-tyk269, v-tyk270, v-tyk271, v-tyk272, v-tyk273, v-tyk274, v-tyk275, v-tyk276, v-tyk277, v-tyk278, v-tyk279, v-tyk280, v-tyk281, v-tyk282, v-tyk283, v-tyk284, v-tyk285, v-tyk286, v-tyk287, v-tyk288, v-tyk289, v-tyk290, v-tyk291, v-tyk292, v-tyk293, v-tyk294, v-tyk295, v-tyk296, v-tyk297, v-tyk298, v-tyk299, v-tyk300, v-tyk301, v-tyk302, v-tyk303, v-tyk304, v-tyk305, v-tyk306, v-tyk307, v-tyk308, v-tyk309, v-tyk310, v-tyk311, v-tyk312, v-tyk313, v-tyk314, v-tyk315, v-tyk316, v-tyk317, v-tyk318, v-tyk319, v-tyk320, v-tyk321, v-tyk322, v-tyk323, v-tyk324, v-tyk325, v-tyk326, v-tyk327, v-tyk328, v-tyk329, v-tyk330, v-tyk331, v-tyk332, v-tyk333, v-tyk334, v-tyk335, v-tyk336, v-tyk337, v-tyk338, v-tyk339, v-tyk340, v-tyk341, v-tyk342, v-tyk343, v-tyk344, v-tyk345, v-tyk346, v-tyk347, v-tyk348, v-tyk349, v-tyk350, v-tyk351, v-tyk352, v-tyk353, v-tyk354, v-tyk355, v-tyk356, v-tyk357, v-tyk358, v-tyk359, v-tyk360, v-tyk361, v-tyk362, v-tyk363, v-tyk364, v-tyk365, v-tyk366, v-tyk367, v-tyk368, v-tyk369, v-tyk370, v-tyk371, v-tyk372, v-tyk373, v-tyk374, v-tyk375, v-tyk376, v-tyk377, v-tyk378, v-tyk379, v-tyk380, v-tyk381, v-tyk382, v-tyk383, v-tyk384, v-tyk385, v-tyk386, v-tyk387, v-tyk388, v-tyk389, v-tyk390, v-tyk391, v-tyk392, v-tyk393, v-tyk394, v-tyk395, v-tyk396, v-tyk397, v-tyk398, v-tyk399, v-tyk400, v-tyk401, v-tyk402, v-tyk403, v-tyk404, v-tyk405, v-tyk406, v-tyk407, v-tyk408, v-tyk409, v-tyk410, v-tyk411, v-tyk412, v-tyk413, v-tyk414, v-tyk415, v-tyk416, v-tyk417, v-tyk418, v-tyk419, v-tyk420, v-tyk421, v-tyk422, v-tyk423, v-tyk424, v-tyk425, v-tyk426, v-tyk427, v-tyk428, v-tyk429, v-tyk430, v-tyk431, v-tyk432, v-tyk433, v-tyk434, v-tyk435, v-tyk436, v-tyk437, v-tyk438, v-tyk439, v-tyk440, v-tyk441, v-tyk442, v-tyk443, v-tyk444, v-tyk445, v-tyk446, v-tyk447, v-tyk448, v-tyk449, v-tyk450, v-tyk451, v-tyk452, v-tyk453, v-tyk454, v-tyk455, v-tyk456, v-tyk457, v-tyk458, v-tyk459, v-tyk460, v-tyk461, v-tyk462, v-tyk463, v-tyk464, v-tyk465, v-tyk466, v-tyk467, v-tyk468, v-tyk469, v-tyk470, v-tyk471, v-tyk472, v-tyk473, v-tyk474, v-tyk475, v-tyk476, v-tyk477, v-tyk478, v-tyk479, v-tyk480, v-tyk481, v-tyk482, v-tyk483, v-tyk484, v-tyk485, v-tyk486, v-tyk487, v-tyk488, v-tyk489, v-tyk490, v-tyk491, v-tyk492, v-tyk493, v-tyk494, v-tyk495, v-tyk496, v-tyk497, v-tyk498, v-tyk499, v-tyk500, v-tyk501, v-tyk502, v-tyk503, v-tyk504, v-tyk505, v-tyk506, v-tyk507, v-tyk508, v-tyk509, v-tyk510, v-tyk511, v-tyk512, v-tyk513, v-tyk514, v-tyk515, v-tyk516, v-tyk517, v-tyk518, v-tyk519, v-tyk520, v-tyk521, v-tyk522, v-tyk523, v-tyk524, v-tyk525, v-tyk526, v-tyk527, v-tyk528, v-tyk529, v-tyk530, v-tyk531, v-tyk532, v-tyk533, v-tyk534, v-tyk535, v-tyk536, v-tyk537, v-tyk538, v-tyk539, v-tyk540, v-tyk541, v-tyk542, v-tyk543, v-tyk544, v-tyk545, v-tyk546, v-tyk547, v-tyk548, v-tyk549, v-tyk550, v-tyk551, v-tyk552, v-tyk553, v-tyk554, v-tyk555, v-tyk556, v-tyk557, v-tyk558, v-tyk559, v-tyk560, v-tyk561, v-tyk562, v-tyk563, v-tyk564, v-tyk565, v-tyk566, v-tyk567, v-tyk568, v-tyk569, v-tyk570, v-tyk571, v-tyk572, v-tyk573, v-tyk574, v-tyk575, v-tyk576, v-tyk577, v-tyk578, v-tyk579, v-tyk580, v-tyk581, v-tyk582, v-tyk583, v-tyk584, v-tyk585, v-tyk586, v-tyk587, v-tyk588, v-tyk589, v-tyk590, v-tyk591, v-tyk592, v-tyk593, v-tyk594, v-tyk595, v-tyk596, v-tyk597, v-tyk598, v-tyk599, v-tyk600, v-tyk601, v-tyk602, v-tyk603, v-tyk604, v-tyk605, v-tyk606, v-tyk607, v-tyk608, v-tyk609, v-tyk610, v-tyk611, v-tyk612, v-tyk613, v-tyk614, v-tyk615, v-tyk616, v-tyk617, v-tyk618, v-tyk619, v-tyk620, v-tyk621, v-tyk622, v-tyk623, v-tyk624, v-tyk625, v-tyk626, v-tyk627, v-tyk628, v-tyk629, v-tyk630, v-tyk631, v-tyk632, v-tyk633, v-tyk634, v-tyk635, v-tyk636, v-tyk637, v-tyk638, v-tyk639, v-tyk640, v-tyk641, v-tyk642, v-tyk643, v-tyk644, v-tyk645, v-tyk646, v-tyk647, v-tyk648, v-tyk649, v-tyk650, v-tyk651, v-tyk652, v-tyk653, v-tyk654, v-tyk655, v-tyk656, v-tyk657, v-tyk658, v-tyk659, v-tyk660, v-tyk661, v-tyk662, v-tyk663, v-tyk664, v-tyk665, v-tyk666, v-tyk667, v-tyk668, v-tyk669, v-tyk670, v-tyk671, v-tyk672, v-tyk673, v-tyk674, v-tyk675, v-tyk676, v-tyk677, v-tyk678, v-tyk679, v-tyk680, v-tyk681, v-tyk682, v-tyk683, v-tyk684, v-tyk685, v-tyk686, v-tyk687, v-tyk688, v-tyk689, v-tyk690, v-tyk691, v-tyk692, v-tyk693, v-tyk694, v-tyk695, v-tyk696, v-tyk697, v-tyk698, v-tyk699, v-tyk700, v-tyk701, v-tyk702, v-tyk703, v-tyk704, v-tyk705, v-tyk706, v-tyk707, v-tyk708, v-tyk709, v-tyk710, v-tyk711, v-tyk712, v-tyk713, v-tyk714, v-tyk715, v-tyk716, v-tyk717, v-tyk718, v-tyk719, v-tyk720, v-tyk721, v-tyk722, v-tyk723, v-tyk724, v-tyk725, v-tyk726, v-tyk727, v-tyk728, v-tyk729, v-tyk730, v-tyk731, v-tyk732, v-tyk733, v-tyk734, v-tyk735, v-tyk736, v-tyk737, v-tyk738, v-tyk739, v-tyk740, v-tyk741, v-tyk742, v-tyk743, v-tyk744, v-tyk745, v-tyk746, v-tyk747, v-tyk748, v-tyk749, v-tyk750, v-tyk751, v-tyk752, v-tyk753, v-tyk754, v-tyk755, v-tyk756, v-tyk757, v-tyk758, v-tyk759, v-tyk760, v-tyk761, v-tyk762, v-tyk763, v-tyk764, v-tyk765, v-tyk766, v-tyk767, v-tyk768, v-tyk769, v-tyk770, v-tyk771, v-tyk772, v-tyk773, v-tyk774, v-tyk775, v-tyk776, v-tyk777, v-tyk778, v-tyk779, v-tyk780, v-tyk781, v-tyk782, v-tyk783, v-tyk784, v-tyk785, v-tyk786, v-tyk787, v-tyk788, v-tyk789, v-tyk790, v-tyk791, v-tyk792, v-tyk793, v-tyk794, v-tyk795, v-tyk796, v-tyk797, v-tyk798, v-tyk799, v-tyk800, v-tyk801, v-tyk802, v-tyk803, v-tyk804, v-tyk805, v-tyk806, v-tyk807, v-tyk808, v-tyk809, v-tyk810, v-tyk811, v-tyk812, v-tyk813, v-tyk814, v-tyk815, v-tyk816, v-tyk817, v-tyk818, v-tyk819, v-tyk820, v-tyk821, v-tyk822, v-tyk823, v-tyk824, v-tyk825, v-tyk826, v-tyk827, v-tyk828, v-tyk829, v-tyk830, v-tyk831, v-tyk832, v-tyk833, v-tyk834, v-tyk835, v-tyk836, v-tyk837, v-tyk838, v-tyk839, v-tyk840, v-tyk841, v-tyk842, v-tyk843, v-tyk844, v-tyk845, v-tyk846, v-tyk847, v-tyk848, v-tyk849, v-tyk850, v-tyk851, v-tyk852, v-tyk853, v-tyk854, v-tyk855, v-tyk856, v-tyk857, v-tyk858, v-tyk859, v-tyk860, v-tyk861, v-tyk862, v-tyk863, v-tyk864, v-tyk865, v-tyk866, v-tyk867, v-tyk868, v-tyk869, v-tyk870, v-tyk871, v-tyk872, v-tyk873, v-tyk874, v-tyk875, v-tyk876, v-tyk877, v-tyk878, v-tyk879, v-tyk880, v-tyk881, v-tyk882, v-tyk883, v-tyk884, v-tyk885, v-tyk886, v-tyk887, v-tyk888, v-tyk889, v-tyk890, v-tyk891, v-tyk892, v-tyk893, v-tyk894, v-tyk895, v-tyk896, v-tyk897, v-tyk898, v-tyk899, v-tyk900, v-tyk901, v-tyk902, v-tyk903, v-tyk904, v-tyk905, v-tyk906, v-tyk907, v-tyk908, v-tyk909, v-tyk910, v-tyk911, v-tyk912, v-tyk913, v-tyk914, v-tyk915, v-tyk916, v-tyk917, v-tyk918, v-tyk919, v-tyk920, v-tyk921, v-tyk922, v-tyk923, v-tyk924, v-tyk925, v-tyk926, v-tyk927, v-tyk928, v-tyk929, v-tyk930, v-tyk931, v-tyk932, v-tyk933, v-tyk934, v-tyk935, v-tyk936, v-tyk937, v-tyk938, v-tyk939, v-tyk940, v-tyk941, v-tyk942, v-tyk943, v-tyk944, v-tyk945, v-tyk946, v-tyk947, v-tyk948, v-tyk949, v-tyk950, v-tyk951, v-tyk952, v-tyk953, v-tyk954, v-tyk955, v-tyk956, v-tyk957, v-tyk958, v-tyk959, v-tyk960, v-tyk961, v-tyk962, v-tyk963, v-tyk964, v-tyk965, v-tyk966, v-tyk967, v-tyk968, v-tyk969, v-tyk970, v-tyk971, v-tyk972, v-tyk973, v-tyk974, v-tyk975, v-tyk976, v-tyk977, v-tyk978, v-tyk979, v-tyk980, v-tyk981, v-tyk982, v-tyk983, v-tyk984, v-tyk985, v-tyk986, v-tyk987, v-tyk988, v-tyk989, v-tyk990, v-tyk991, v-tyk992, v-tyk993, v-tyk994, v-tyk995, v-tyk996, v-tyk997, v-tyk998, v-tyk999, v-tyk1000, v-tyk1001, v-tyk1002, v-tyk1003, v-tyk1004, v-tyk1005, v-tyk1006, v-tyk1007, v-tyk1008, v-tyk1009, v-tyk1010, v-tyk1011, v-tyk1012, v-tyk1013, v-tyk1014, v-tyk1015, v-tyk1016, v-tyk1017, v-tyk1018, v-tyk1019, v-tyk1020, v-tyk1021, v-tyk1022, v-tyk1023, v-tyk1024, v-tyk1025, v-tyk1026, v-tyk1027, v-tyk1028, v-tyk1029, v-tyk1030, v-tyk1031, v-tyk1032, v-tyk1033, v-tyk1034, v-tyk1035, v-tyk1036, v-tyk1037, v-tyk1038, v-tyk1039, v-tyk1040, v-tyk1041, v-tyk1042, v-tyk1043, v-tyk1044, v-tyk1045, v-tyk1046, v-tyk1047, v-tyk1048, v-tyk1049, v-tyk1050, v-tyk1051, v-tyk1052, v-tyk1053, v-tyk1054, v-tyk1055, v-tyk1056, v-tyk1057, v-tyk1058, v-tyk1059, v-tyk1060, v-tyk1061, v-tyk1062, v-tyk1063, v-tyk1064, v-tyk1065, v-tyk1066, v-tyk1067, v-tyk1068, v-tyk1069, v-tyk1070, v-tyk1071, v-tyk1072, v-tyk1073, v-tyk1074, v-tyk1075, v-tyk1076, v-tyk1077, v-tyk1078, v-tyk1079, v-tyk1080, v-tyk1081, v-tyk1082, v-tyk1083, v-tyk1084, v-tyk1085, v-tyk1086, v-tyk1087, v-tyk1088, v-tyk1089, v-tyk1090, v-tyk1091, v-tyk1092, v-tyk1093, v-tyk1094, v-tyk1095, v-tyk1096, v-tyk1097, v-tyk1098, v-tyk1099, v-tyk1100, v-tyk1101, v-tyk1102, v-tyk1103, v-tyk1104, v-tyk1105, v-tyk1106, v-tyk1107, v-tyk1108, v-tyk1109, v-tyk1110, v-tyk1111, v-tyk1112, v-tyk1113, v-tyk1114, v-tyk1115, v-tyk1116, v-tyk1117, v-tyk1118, v-tyk1119, v-tyk1120, v-tyk1121, v-tyk1122, v-tyk1123, v-tyk1124, v-tyk1125, v-tyk1126, v-tyk1127, v-tyk1128, v-tyk1129, v-tyk1130, v-tyk1131, v-tyk1132, v-tyk1133, v-tyk1134, v-tyk1135, v-tyk1136, v-tyk1137, v-tyk1138, v-tyk1139, v-tyk1140, v-tyk1141, v-tyk1142, v-tyk1143, v-tyk1144, v-tyk1145, v-tyk1146, v-tyk1147, v-tyk1148, v-tyk1149, v-tyk1150, v-tyk1151, v-tyk1152, v-tyk1153, v-tyk1154, v-tyk1155, v-tyk1156, v-tyk1157, v-tyk1158, v-tyk1159, v-tyk1160, v-tyk1161, v-tyk1162, v-tyk1163, v-tyk1164, v-tyk1165, v-tyk1166, v-tyk1167, v-tyk1168, v-tyk1169, v-tyk1170, v-tyk1171, v-tyk1172, v-tyk1173, v-tyk1174, v-tyk1175, v-tyk1176, v-tyk1177, v-tyk1178, v-tyk1179, v-tyk1180, v-tyk1181, v-tyk1182, v-tyk1183, v-tyk1184, v-tyk1185, v-tyk1186, v-tyk1187, v-tyk1188, v-tyk1189, v-tyk1190, v-tyk1191, v-tyk1192, v-tyk1193, v-tyk1194, v-tyk1195, v-tyk1196, v-tyk1197, v-tyk1198, v-tyk1199, v-tyk1200, v-tyk1201, v-tyk1202, v-tyk1203, v-tyk1204, v-tyk1205, v-tyk1206, v-tyk1207, v-tyk1208, v-tyk1209, v-tyk1210, v-tyk1211, v-tyk1212, v-tyk1213, v-tyk1214, v-tyk1215, v-tyk1216, v-tyk1217, v-tyk1218, v-tyk1219, v-tyk1220, v-tyk1221, v-tyk1222, v-tyk1223, v-tyk1224, v-tyk1225, v-tyk1226, v-tyk1227, v-tyk1228, v-tyk1229, v-tyk1230, v-tyk1231, v-tyk1232, v-tyk1233, v-tyk1234, v-tyk1235, v-tyk1236, v-tyk1237, v-tyk1238, v-tyk1239, v-tyk1240, v-tyk1241, v-tyk1242, v-tyk1243, v-tyk1244, v-tyk1245, v-tyk1246, v-tyk1247, v-tyk1248, v-tyk1249, v-tyk1250, v-tyk1251, v-tyk1252, v-tyk1253, v-tyk1254, v-tyk1255, v-tyk1256, v-tyk1257, v-tyk1258, v-tyk1259, v-tyk1260, v-tyk1261, v-tyk1262, v-tyk1263, v-tyk1264, v-tyk1265, v-tyk1266, v-tyk1267, v-tyk1268, v-tyk1269, v-tyk1270, v-tyk1271, v-tyk1272, v-tyk1273, v-tyk1274, v-tyk1275, v-tyk1276, v-tyk1277, v-tyk1278, v-tyk1279, v-tyk1280, v-tyk1281, v-tyk1282, v-tyk1283, v-tyk1284, v-tyk1285, v-tyk1286, v-tyk1287, v-tyk1288, v-tyk1289, v-tyk1290, v-tyk1291, v-tyk1292, v-tyk1293, v-tyk1294, v-tyk1295, v-tyk1296, v-tyk1297, v-tyk1298, v-tyk1299, v-tyk1300, v-tyk1301, v-tyk1302, v-tyk1303, v-tyk1304, v-tyk1305, v-tyk1306, v-tyk1307, v-tyk1308, v-tyk1309, v-tyk1310, v-tyk1311, v-tyk1312, v-tyk1313, v-tyk1314, v-tyk1315, v-tyk1316, v-tyk1317, v-tyk1318, v-tyk1319, v-tyk1320, v-tyk1321, v-tyk1322, v-tyk1323, v-tyk1324, v-tyk1325, v-tyk1326, v-tyk1327, v-tyk1328, v-tyk1329, v-tyk1330, v-tyk1331, v-tyk1332, v-tyk1333, v-tyk1334, v-tyk1335, v-tyk1336, v-tyk1337, v-tyk1338, v-tyk1339, v-tyk1340, v-tyk1341, v-tyk1342, v-tyk1343, v-tyk1344, v-tyk1345, v-tyk1346, v-tyk1347, v-tyk1348, v-tyk1349, v-tyk1350, v-tyk1351, v-tyk1352, v-tyk1353, v-tyk1354, v-tyk1355, v-tyk1356, v-tyk1357, v-tyk1358, v-tyk1359, v-tyk1360, v-tyk1361, v-tyk1362, v-tyk1363, v-tyk1364, v-tyk1365, v-tyk1366, v-tyk1367, v-tyk1368, v-tyk1369, v-tyk1370, v-tyk1371, v-tyk1372, v-tyk1373, v-tyk1374, v-tyk1375, v-tyk1376, v-tyk1377



```

QY 420 CAG---TCCAGGAGCCAGTAAATGA-----GAGCCCCAAAAAGAACACGCA 467
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 227 GluGlySerProAlaGlyGlyGlyGluAlaGluProGlnAspSerProSerAla 246

QY 468 GCTGAAGTCGGACTCCTACACCTGGGAGCAGACAGAGAAGATCAGATACAGTGA 527
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 247 AlaSerGlnGluAspProGluSerAspValSerAspAspSerAspGlnGluVal----- 264

QY 528 ATCCAGTCGGCAGACATGAAGTGAT 554
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 265 -----AspIleGluGlyAsp 269

RESULT 14
B30843
glutenin high molecular weight chain Ax2 precursor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 15-Nov-1996
C:Accession: B30843
R:Anderson, O.D.; Greene, F.C.
submitted to GenBank, January 1989
A:Reference number: A94515
A:Accession: B30843
A:Molecule type: DNA
A:Residues: 1-815 <AND>
C:Superfamily: glutenin
C:Keywords: seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-815/Product: glutenin Ax2 chain #status predicted <MAT>
F:108-773/Region: glutamine/glycine/proline-rich

Alignment Scores:
Pred. No.: 2.53 Length: 815
Score: 96.50 Matches: 45
Percent Similarity: 33.01% Conservative: 23
Best Local Similarity: 21.84% Mismatches: 73
Query Match: 6.99% Indels: 66
DB: 2 Gaps: 6

US-09-658-824-808 (1-781) x B30843 (1-815)
QY 61 GAGACACGGCGGTAGTCCACAGGAGTCCCACTGGGAGTTGAAGTGTGAGTGAGT 120
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 50 GlnLeuArgAspValSerProGluCysGlnProValGlyGlyGlyProValAlaArgGln 69

QY 121 GAAGAGAACACAGAGCTTCCGAGGGTTGTGTGTGTCAGTCACTCAGAGTGAGAGGCC 180
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 70 TyrGluGln----- 72

QY 181 CTCGAAGTCGTCTCCTCTCATCGGTGCCAGCCCATGGACCTTCTTGTCTGTCAAG 240
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 73 ---GlnValValPro----- 77

QY 241 GCCATNACTAGGAGGAAGAGGCGCGAGAGTGGAGGG----- 279
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 78 -----ProLysGlyGlySerPheTyrProGlyGluThrThr 89

QY 280 -----GCTCAGCGAAGCTGGGTGCTGTGGGGGTATCCGAGT----- 318
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 90 ProProGlnGlnLeuGlnGlnSerIleLeuTrpGlyIleProAlaLeuLeuArgArgTyr 109

QY 319 CCCAGAGAACCTTGAACCCCGACAGAGATTTCTGCATCTCC----- 366
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 110 TyrLeuSerValThrSerProGlnGlnValSerTyrTyrProGlyGlnAlaSerSerGln 129

QY 367 GGACAGGAGGAGGCGGATGAGCGACAC-ACAAACACAGAACACACAGCCAGTCC 425
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 130 ArgProGlyGlnGlyGlnGlnGluTyrTyrLeuThrSerProGlnGlnSerGlyGlnTrp 149

QY 426 CAGGAGCCAGTAATGGAGAG-----CCCCAAAAAGAACACCA 464
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 150 GlnGlnProGlyGlnGlnSerGlyTyrTyrProThrSerProGlnGlnSerGlyGln 169

QY 465 GCAGCTGAAGTCGGATCCTACACCTGGCAGCAGACAGAGAAGATCAGGATACAGCT 524

```

```

Db 170 LysGlnProGlyTyrTyrProThrSerProThrProGlnProGlnLeuGlnProThr 189
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 525 GAGATCCCACTGCGGCACATGGAAGTGTATCGCAAGAGCTGCATCAGTCAACACCCGG 584
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 190 GlnGlyGlnArgGlnGlnProGlyGlnGlyGlnGlnLeuArgGlnGlyGlnGlnGly 209

QY 585 GATAAATCTGGATTGGG 602
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 210 GlnGlnSerGlyGlnGly 215

RESULT 15
JN0689
glutenin, high-molecular-weight Ax2* chain precursor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 15-Nov-1996
C:Accession: JN0689
R:Anderson, O.D.; Greene, F.C.
Theor. Appl. Genet. 77, 689-700, 1989
A:Title: The characterization and comparative analysis of high-molecular-weight glute
A:Reference number: JN0689
A:Accession: JN0689
A:Molecule type: DNA
A:Residues: 1-815 <AND>
C:Superfamily: glutenin
C:Keywords: seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-815/Product: glutenin, high-molecular-weight Ax2* chain #status predicted <MAT>
F:108-773/Region: glutamine/glycine/proline-rich

Alignment Scores:
Pred. No.: 2.53 Length: 815
Score: 96.50 Matches: 45
Percent Similarity: 33.01% Conservative: 23
Best Local Similarity: 21.84% Mismatches: 73
Query Match: 6.99% Indels: 66
DB: 2 Gaps: 6

US-09-658-824-808 (1-781) x JN0689 (1-815)
QY 61 GAGACACGGCGGTAGTCCACAGGAGTCCCACTGGGAGTTGAAGTGTGAGTGAGT 120
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 50 GlnLeuArgAspValSerProGluCysGlnProValGlyGlyGlyProValAlaArgGln 69

QY 121 GAAGAGAACACAGAGCTTCCGAGGGTTGTGTGTGTCAGTCACTCAGAGTGAGAGGCC 180
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 70 TyrGluGln----- 72

QY 181 CTCGAAGTCGTCTCCTCTCATCGGTGCCAGCCCATGGACCTTCTTGTCTGTCAAG 240
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 73 ---GlnValValPro----- 77

QY 241 GCCATNACTAGGAGGAAGAGGCGCGAGAGTGGAGGG----- 279
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 78 -----ProLysGlyGlySerPheTyrProGlyGluThrThr 89

QY 280 -----GCTCAGCGAAGCTGGGTGCTGTGGGGGTATCCGAGT----- 318
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 90 ProProGlnGlnLeuGlnGlnSerIleLeuTrpGlyIleProAlaLeuLeuArgArgTyr 109

QY 319 CCCAGAGAACCTTGAACCCCGACAGAGATTTCTGCATCTCC----- 366
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 110 TyrLeuSerValThrSerProGlnGlnValSerTyrTyrProGlyGlnAlaSerSerGln 129

QY 367 GGACAGGAGGAGGCGGATGAGCGACAC-ACAAACACAGAACACACAGCCAGTCC 425
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 130 ArgProGlyGlnGlyGlnGlnGluTyrTyrLeuThrSerProGlnGlnSerGlyGlnTrp 149

QY 426 CAGGAGCCAGTAATGGAGAG-----CCCCAAAAAGAACACCA 464
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```







A:Reference number: S10001; MUID:90251472

C:Accession: S10001

A:Molecule type: mRNA

A:Residues: 1-657 <LOR>

A:Cross-references: EMBL:X51829; NID:953040; PIDN:CAA36128.1; PID:953041

Alignment Scores:

Pred. No.: 3.36 Length: 657  
Score: 95.00 Matches: 61  
Percent Similarity: 35.77% Conservative: 27  
Best Local Similarity: 24.80% Mismatches: 93  
Query Match: 6.88% Indels: 65  
DB: 2 Gaps: 12

US-09-658-824-808 (1-781) x S10001 (1-657)

```
QY 55 GCTACTGACACGGCGGTAGGTCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGT 114
Db 212 AlaThrGluGluGlyThrGluAsnLysAlaAspProSerAsnSerProSer----- 229
QY 115 GAGAGTGAAGGAACACAGCAGGCTCCGGAGGTGTGTGTCAGTCACTCAGAGTCAG 174
Db 230 -----SerGlySerHisSerArgAlaTrpGluTyrTyrSerArgGlu 243
QY 175 AAGGGCTCTCGAAGTCGTGCTCCCTCTCATGGGTGCCCGCCATGGACCTTCTTGTCTC 234
Db 244 LysProLysGlnGluGlyGluAlaLysValGluAlaHisArgAlaGlyGlnGlyHisPro 263
QY 235 GTCAGCGCCATACTAGGAGGAGGAGGCGCGAGG-----AGTGGAGGG 279
Db 264 CysArgAsnAlaGluAlaGluGluGlyGlyProGluThrPheValCysThrGlyAsn 283
QY 280 GCTCAGGGGAAGCTGGGTGTGTGGGGGTATCCGAGTCCAGACGACCTGGAAACCCC 339
Db 284 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 296
QY 340 GACAGAGATTCGTGACTCC-----CCAGACGGGACCAGAGAGGAGCGCATGAGCGA 393
Db 296 pThrGluGluGlu-AspAsnSerAspSerAspSerAlaGluGluAspThrAla----- 313
QY 394 CACACAAACACACAAACACACAGCCAGTCC----- 425
Db 314 --GlnThrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 333
QY 426 -----CAGGAGCCCAATAATGGAGCCCAAAAGAGAACACAGCAGCTG 471
Db 333 lylGluAspThrGluGluAspSerAspSerAspSerAlaGluGluAspThrAlaGlnT 353
QY 472 AAAGTCGGGATCCT-----ACACCTGGGCAGC 498
Db 353 hrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyGluA 373
QY 499 AGACAGAGAAGATCAGGATACAGCTGAGATCCCACTGCGGCACATGGAAGGTGATCTGC 558
Db 373 spThrGluGluGluAsnSer-----AspLeuAspSerAlaGluG 386
QY 559 AAGAGCTGCATCAGTCAAC-----ACCGGGGATAATCTGGATT-----GGGTCC 606
Db 386 lAspThrAlaGlnThrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValT 406
QY 607 GCGGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAATGCCAAGACGAGTCAA 666
Db 406 yrArgProGlyGluAsp---Thr-GluGluGluAsnSerAspLeuAspSerAlaGluGlu 424
QY 667 GAGCAACCCAA 678
Db 425 AspThrAlaGln 428
```

RESULT 19

AF2074

ferrichrome-iron receptor [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002

C:Accession: AF2074

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AF2074

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-863 <LOR>

A:Cross-references: GB:BA000019; PIDN:BA073847.1; PID:gl7131239; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all2148

Alignment Scores:  
Pred. No.: 3.79 Length: 863  
Score: 94.50 Matches: 53  
Percent Similarity: 45.13% Conservative: 35  
Best Local Similarity: 27.18% Mismatches: 66  
Query Match: 6.85% Indels: 42  
DB: 2 Gaps: 11

US-09-658-824-808 (1-781) x AF2074 (1-863)

```
QY 115 GAGAGTGAAGAG-----GAACCAGCAGGCTCCGAGGGTTGTGTGTCAGTCACTCAG 168
Db 58 AspAlaGluGluLeuValGluValThrGlyValArg-----IleAsnGln 72
QY 169 AGTGAGAAGGCCCTCGAAGTCGTGCTCCCTCTCATGCGGTGCCACGCCATGGACCT--- 225
Db 73 ThrGluLysGlyValGluValIleLeuGluThrAlaLysGlyAspAlaLeuLysProVal 92
QY 226 -----TCTTGTCTGTCACGGCCATA-----ACTAGGAGAGAA 258
Db 93 GlnLysAsnGluGlyAsnLysLeuIleLeuAspIleProAsnSerGlnLeuArgTyrGlu 112
QY 259 GGAGGGCC-----GAGGAGTGGAGGGCTCAGCGGAAGCTGGGTGCTGTTGGGG 308
Db 113 GlyGlyAspThrPheArgGlnGluLysProPheAlaGlyIleAlaGluValLeuVal--- 131
QY 309 GTATCCGAGTCCAGAGACCTCGAA-----CCCCGA 341
Db 132 ValAsnGlnAspAsnAsnThrIleGlnValThrValThrGlyGluThrGlyLeuProIle 151
QY 342 CAGAAGATTCTGGACTCCCAAGAGGGGACCAAGAGAGGAGCGGATGAGCGCACACACA 401
Db 152 ValGluLeuPheAspGlyAspGlyLeuIlePheGlyValThrProAlaThrThrThr 171
QY 402 AACACAGAACACACAGC---CAGTCCAGGAGCCCAAGTAATGGAGAGCCCAAAAGAA 458
Db 172 AlaGlnGlnProGlnThrProGlnAlaGlnGluLysProAlaIleGluIleProGlnGlu 191
QY 459 GAACACGAGCTGAAAGTCGGGATCCT-----ACACCTGGGCACACACA--- 503
Db 192 GluProAlaAlaGlnGlnAspGluProIleGluLeuValValThrGlyGlnGlnAsnGly 211
QY 504 -----GAAGAAGATCAGGATACAGCT-----GAGATCCAGTCGCGGCACATCGAA 548
Db 212 TyrArgValGlnAspAlaThrThrAlaThrLysThrAspThrProLeuArgAspIlePro 231
QY 549 GGTGATCTGCAAGAGCTGCATCAGTCAACACACCGGGGATAAATCT 593
Db 232 GlnSerIleGlnValProArgGluValLeuGluAspArgAsn 246
```

RESULT 20

S55659

tegument protein 64 - equine herpesvirus 2

C:Species: equine herpesvirus 2

C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Oct-1999

C:Accession: S55659

R:Telford, E.A.R.; Watson, M.S.; Alrd, H.C.; Perry, J.; Davison, A.J.

J. Mol. Biol. 249, 520-528, 1995

A: Title: The DNA sequence of equine herpesvirus 2.

A: Reference number: S55594; MOID: 95502501

A: Accession: S55659

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-3436 <TEL>

A: Cross-references: GB:U020824; MID: g965172; PIDN: AAC13852.1; PID: g6955237

A: Note: The nucleotide sequence was submitted to the EMBL Data Library, February 1995

A: Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 99

Alignment Scores: 4.24 Length: 1032  
Pred. No.: 94.00 Matches: 72  
Score: 33.70% Conservative: 19  
Percent Similarity: 26.67% Mismatches: 101  
Best Local Similarity: 6.59% Indels: 78  
Query Match: 2 Gaps: 10  
DB:

US-09-658-824-808 (1-781) x T34433 (1-1032)

```
QY 604 AACCCAAATCCAGATTATCCCGGTGTGACTGATGCAGCTCTTGGAGATCACCTTCC 545
   ::::::::::::::::::::
Db 534 SerProAsnProSerSerSerProAlaSerThrGlySerThrIleThrIleSerGlySer 553
QY 544 ATGTCG-----CGCACTGGGATCTCA 524
   |||
Db 554 SerSerIleValSerThrValSerGlySerThrValSerGlySerThrGlyThrSer 573
QY 523 GCTGATCTCTTCTTCTGCTGCTGCCAGGTGTAGGATCCCGACTTTTCAGCTGCT 464
   ::::::::::::::::::::
Db 574 GlnSerThrLeuAlaSerSerThrAlaThrProGlySerSerSerThrValProSerSer 593
QY 463 GGTCTCTCTTTTGGGGCTCTCATTACTGGCTCTGGGACTGGCTGTGGTCTGTG 404
   |||
Db 594 SerSerProGlnProSerSer-GlnSerProAlaProAsnThrGlySerThrThrProse 613
QY 403 TTTGTGTGCTGCTCTATCGCTCCCTCTCTCTGG----- 370
   |||
Db 613 rGlnThrSerSerGlnSerProSerProSerMetAsnProSerSerSerThrProThrG 633
QY 369 -----TCCCGTCTGGGAGTCCAGATCTTCTCT 341
   |||
Db 633 ySerSerGlnSerThrIleThrProGluGlySerThrAlaSerSerProThrGlySerTh 653
QY 340 CGGG-----GTTCCAGGTCTTC 323
   |||
Db 653 rGlySerThrPheSerValAlaThrGluValThrSerGlnSerThrValProSerGlySe 673
QY 322 T-----GGGACTCGGATACCCCAACAGACCCACAG-----CCTTCGCTCAGCCCTCCAC 272
   ::::::::::::::::::::
Db 733 erSerThrSerGlyAlaThrThrSerSerGlySerAlaGlyThrThrMetThrSerProS 753
QY 133 CTGTTCTCTTCTACTCTCCTCACTTCACTCCAGTTGAGTGTGCTGTGGACCTA 74
   ::::::::::::::::::::
Db 753 erGlnSerSer-----ValGlySerSerGlnGlySert 765
QY 73 CCGGC---CGTGCTCTAGTACGGAGAGAAATCCAGACCTCAGGA-----C 29
   |||
Db 765 hrSerProAlaAlaSerThrSerGlyGluMetThrSerGlnGlySerThrGlnThrP 785
QY 28 CCGAGTCGCGGCTCAGCTTCCGCTCC 3
   |||
Db 785 roGlySerSerValSerThrSerAla 793
```

RESULT 23

T34434

hypothetical protein K06A9.1a - Caenorhabditis elegans

C: Species: Caenorhabditis elegans

C: Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000

C: Accession: T34434  
R: Geisler, C.; Gattung, S.  
submitted to the EMBL Data Library, December 1996  
A: Description: The sequence of C. elegans cosmid K06A9.  
A: Reference number: 221525  
A: Accession: T34434  
A: Status: preliminary; translated from GB/EMBL/DBJ  
A: Molecule type: DNA  
A: Residues: 1-2232 <GEI>  
A: Cross-references: EMBL:U08046; PIDN: AAC70890.1; GSPDB: GN00028; CESP: K06A9.1a  
A: Experimental source: strain Bristol N2; clone K06A9  
C: Genetics:  
A: Gene: CESP: K06A9.1a  
A: Map position: X  
A: Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2

Alignment Scores: 4.48 Length: 2232  
Pred. No.: 94.00 Matches: 72  
Score: 33.70% Conservative: 19  
Percent Similarity: 26.67% Mismatches: 101  
Best Local Similarity: 6.59% Indels: 78  
Query Match: 2 Gaps: 10  
DB:

US-09-658-824-808 (1-781) x T34434 (1-2232)

```
QY 604 AACCCAAATCCAGATTATCCCGGTGTGACTGATGCAGCTCTTGGAGATCACCTTCC 545
   ::::::::::::::::::::
Db 534 SerProAsnProSerSerSerProAlaSerThrGlySerThrIleThrIleSerGlySer 553
QY 544 ATGTCG-----CGCACTGGGATCTCA 524
   |||
Db 554 SerSerIleValSerThrValSerGlySerThrValSerGlySerThrGlyThrSer 573
QY 523 GCTGATCTCTTCTTCTGCTGCTGCCAGGTGTAGGATCCCGACTTTTCAGCTGCT 464
   ::::::::::::::::::::
Db 574 GlnSerThrLeuAlaSerSerThrAlaThrProGlySerSerSerThrValProSerSer 593
QY 463 GGTCTCTCTTTTGGGGCTCTCATTACTGGCTCTGGGACTGGCTGTGGTCTGTG 404
   |||
Db 594 SerSerProGlnProSerSer-GlnSerProAlaProAsnThrGlySerThrThrProse 613
QY 403 TTTGTGTGCTGCTCTATCGCTCCCTCTCTCTGG----- 370
   |||
Db 613 rGlnThrSerSerGlnSerProSerProSerMetAsnProSerSerSerThrProThrG 633
QY 369 -----TCCCGTCTGGGAGTCCAGATCTTCTCT 341
   |||
Db 633 ySerSerGlnSerThrIleThrProGluGlySerThrAlaSerSerProThrGlySerTh 653
QY 340 CGGG-----GTTCCAGGTCTTC 323
   |||
Db 653 rGlySerThrPheSerValAlaThrGluValThrSerGlnSerThrValProSerGlySe 673
QY 322 T-----GGGACTCGGATACCCCAACAGACCCACAG-----CCTTCGCTCAGCCCTCCAC 272
   ::::::::::::::::::::
Db 673 rSerLeuGlyThrGlnSerThrAsnSerProSerProSerSerSerLeuSerProSerTh 693
QY 271 T-----CCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 239
   |||
Db 693 rSerGlyMetSerThrLeuThrSerGluProSerProSerSerThrGlnSerSerGlyAl 713
QY 238 TGACGAGACAAGAGTCCATGGCGCTGGCCGATGAGAGGAGC----- 192
   |||
Db 713 aGlnSerThrLeuThrProSerProAsnProSerGln-SerThrSerSerSerLeuGlu 733
QY 191 --ACGACTTCGAGGCTCTCTACTCTGAGTCACTACCACACACCCCTCCGGAAGCCTG 134
   ::::::::::::::::::::
Db 733 erSerThrSerGlyAlaThrThrSerSerGlySerAlaGlyThrThrMetThrSerProS 753
QY 133 CTGTTCTCTTCTACTCTCCTCACTTCACTCCAGTTGAGTGTGCTGTGGACCTA 74
   ::::::::::::::::::::
Db 753 erGlnSerSer-----ValGlySerSerGlnGlySert 765
```





A: Introns: 42/3; 108/3; 141/3; 210/3; 232/3; 501/1  
C: Superfamily: Caenorhabditis elegans hypothetical protein Y7588A.13

Alignment Scores:  
Pred. No.: 5,47 Length: 540  
Score: 92.50 Matches: 40  
Percent Similarity: 40.85% Conservatives: 18  
Best Local Similarity: 28.17% Mismatches: 56  
Query Match: 6.70% Indels: 28  
DB: 2 Gaps: 8

US-09-658-824-808 (1-781) x T27400 (1-540)

QY 279 GGCTCAGCGAGCTGGG---GTGCTGTTGGGGTATCCAGTCCCGAAGACACCTGGAA 335  
Db 310 GYAlaGlyGluAlaGlySerIleThrValThrAlaSerSerSerGluIys----- 326  
QY 336 CCCCAGCAGAAAGATTCTGGACTCCCGACAGCGGACAGGAGCGGATGAGCGACA 395  
Db 327 -----SerPro---AlaAspSerGlnGlySerAla---SerPro 337  
QY 396 CACACAACACACACACACACCGCAGTCCCGAGGCGCCAGTAAT----- 440  
Db 338 AspGluSerThrGluGluProSerGlnAspHisThrProGlnAspLeuProMetAspGln 357  
QY 441 GGAGAGCCCCAAAAGAAAGACAGCAGCTGAAAGTCGGGATCCTACACCTGGCGACGAG 500  
Db 358 AlaProGlnAlaGluAspLeuGluArgSerAspAlaAspProGlySerArg 377  
QY 501 ACAGAGACAGTACAGTACAGTCCCGAGTCCCGAGCAGTGAAGGTGATCTGCAA 560  
Db 378 AspGluSerGlnAsnAspVal-----ThrAlaSerLeuAspGlySerThrGlu 393  
QY 561 GAGTGCATCATCAACACCC-----GGGGATAAATCTGGATTGGTTCGCGCGT 611  
Db 394 GluAlaSerGlnAspHisThrGluPheValAlaAspThrProGlyAlaGlyAspGluSer 413  
QY 612 CAGGTGAA-----GATTAATCTAAAGAGGAACACTGTAAATGCCAGACAGGTGA 665  
Db 414 ThrAspAspAlaLeuAsn-LeuProMetAspThrAspGlnAlaProProHisAlaGlyG 433  
QY 666 AGAG 669  
Db 433 uGlu 434

RESULT 29  
A34615  
C: Species: Rattus norvegicus (Norway rat)  
C: Date: 29-Jun-1990 #sequence\_revision 09-Oct-1992 #text\_change 04-Feb-2000  
C: Accession: A34615  
R: Haydock, P.V.; Dale, B.A.  
A: Title: Filaggrin, an intermediate filament-associated protein: structural and function  
A: Reference number: A34615; MUID: 90274870  
A: Accession: A34615  
A: Status: Preliminary  
A: Molecule type: mRNA  
A: Residues: 1-625 <HAY>  
A: Cross-references: GB:M21759; NID:g204143; PIDN:AAA41161.1; PID:g204144  
A: Note: The authors translated the codon GAA for residue 568 as Gln  
C: Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology  
C: Keywords: EF hand; epidermis

Alignment Scores:  
Pred. No.: 5,53 Length: 625  
Score: 92.50 Matches: 48  
Percent Similarity: 34.22% Conservatives: 16  
Best Local Similarity: 25.67% Mismatches: 82  
Query Match: 6.70% Indels: 41  
DB: 2 Gaps: 9

US-09-658-824-808 (1-781) x A34615 (1-625)

QY 141 CCGAGGGTTGTGTGGTCACTGACTCAGAGCTGAGAGGCCCTCGAAGTCCTGCTCCCTCT 200  
Db 9 ProGlnArgGluAlaAlaAlaArgAspSerSer---GluGlnAlaGlnSerArgArg---Thr 26  
QY 201 CATGCGGTGCCACGCCCATGGACCTTCTGTCGTACGCGCCATCACTAGGAGGAGG 260  
Db 27 GluThrIleSerArg-----GlyArgSerGlyHisSerThrGly----- 39  
QY 261 AGGCGCCAGCAGTGGAGG-----GGCTCAGCGCGAAGCT 293  
Db 40 ArgAlaHisGluAspArgHisGluGlnAlaThrAspArgSerAlaGlySerGlySerArg 59  
QY 294 GGGTCTCTGTGGGGTATCCAGTCCCGAAGCACCCTGGAAACCCCGACAGAGATTCTG 353  
Db 60 GlyGlyGlnAlaGly---SerHisSerGluSerGluAlaSerGlyGlyGln----- 75  
QY 354 GACTCCCCAGACGGGACGAGGAGGAGGAGGATGACGACACACACACAGAACCA 413  
Db 76 -----AlaGlyArgArgGlyThrAlaAlaThrArgHisThrSerArgProGlu 91  
QY 414 CACAGCCAGTCCCGAGGCGCCAGTAATGGAGAGCCCCAAAGAAAGAACACGAGCTGAA 473  
Db 92 GlnSerProAspThrAlaGlyArgThrGlySerSerArgGlyGlnGlnSerAlaGlnArg 111  
QY 474 AGTCGGGATCTACACCTGGCGACGACAGACAGAAAGATCAGGATACAGCTGAGATCCCA 533  
Db 112 HisGlyAspSerThrProGlySerThrArgThrGlySerArgGlyArgGlyGluSerPro 131  
QY 534 -----GTCCGCGACATGGAAGGTGATCTGCAA 560  
Db 132 AlaGlyGlnGlnSerProAspArgAlaArgHisIleGluSerArgArgGlyArgThrArg 151  
QY 561 GAGTGCATCATCAACACCCGGGATAAATCTGGATTTGGTTCGCGCGT----- 611  
Db 152 GluAlaSerAlaSerGlnSerSerAspSerGluGlyHisSerGlyAlaHisAlaGlyIle 171  
QY 612 ---CAGGTGAAAGTAATACC 629  
Db 172 GlyGlnGlyGlnThrSerThr 178

RESULT 30  
QRHUT1  
Microtubule-associated protein tau, long splice form - human  
N: Alternate names: microtubule-binding protein tau; neurofibrillary tangle protein tau  
N: Contains: microtubule-associated protein tau type II; microtubule-associated protein tau  
C: Species: Homo sapiens (man)  
C: Date: 30-Jun-1990 #sequence\_revision 03-May-1996 #text\_change 22-Jun-1999  
C: Accession: J50370; A30217; JN0009; S03796; S26665; S26666; S26662; S17302; A43444;  
R: Goedert, M.; Spillantini, M.G.; Jakes, R.; Rutherford, D.; Crowther, R.A.  
Neuron 3, 519-526, 1989  
A: Title: Multiple isoforms of human microtubule-associated protein tau: sequences ar  
A: Reference number: J50370; MUID: 90380393  
A: Accession: J50370  
A: Molecule type: mRNA  
A: Residues: 1-441 <GOE>  
A: Note: six isoforms are found; the clone httau40 sequence is shown. Residues 45-73, the clone httau24 sequence lacks inserts 1 and 2; the clone httau37 sequence lacks ir  
R: Goedert, M.; Wischik, C.M.; Crowther, R.A.; Walker, J.E.; Klug, A.  
Proc. Natl. Acad. Sci. U.S.A. 85, 4051-4055, 1988  
A: Title: Cloning and sequencing of the cDNA encoding a core protein of the paired he  
A: Reference number: A30217; MUID: 88234557  
A: Accession: A30217  
A: Molecule type: mRNA  
A: Residues: 1-44,103-274,306-441 <GO2>  
A: Cross-references: GB:J03778; NID:g338684; PIDN:AAA60615.1; PID:g338685  
R: Lee, G.; Neve, R.L.; Kosik, R.S.  
Neuron 2, 1615-1624, 1989  
A: Title: The microtubule binding domain of tau protein.  
A: Reference number: JN0009; MUID: 90180482  
A: Accession: JN0009  
A: Molecule type: mRNA  
A: Residues: 1-44,103-274,306-441 <LEE>

**Alignment Scores:**







A35962  
protein-tyrosine kinase (EC 2.7.1.112) I, splice form B - human  
N:Alternate names: c-abl proto-oncogene homolog 2  
C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 04-Feb-2000  
R:Kruh, G.D.; Perego, R.; Miki, T.; Aaronson, S.A.  
Proc. Natl. Acad. Sci. U.S.A. 87, 5802-5806, 1990  
A:Title: The complete coding sequence of arg defines the Abelson subfamily of cytoplasmic tyrosine kinases  
A:Reference number: A35962; MUID:90332670  
A:Accession: A35962  
A:Molecule type: mRNA  
A:Residues: 1-1182 <KRUI>  
A:CROSS-references: GB:M35296; NID:g178992; PID:AAA35553.1; PID:g178993  
R:Kruh, G.D.; King, C.R.; Kraus, M.H.; Popescu, N.C.; Amsbaugh, S.C.; McBride, W.O.; Aar  
Science 234, 1545-1548, 1986  
A:Title: A novel human gene closely related to the abl proto-oncogene.  
A:Reference number: A47577; MUID:87069952  
A:Accession: A47577  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 'T', 345-469 <KRUI>  
A:CROSS-references: GB:M14903; NID:g177954; GB:M14904; NID:g177955  
C:Genetics:  
A:Gene: GDB:ABL2; ABL1  
A:CROSS-references: GDB:119641; OMIM:164690  
A:Map position: 1q24-lq25  
C:Superfamily: human protein-tyrosine kinase abl; protein kinase homology; SH2 homology;  
F:114-162/Domain: SH3 homology <SH3>  
F:173-263/Domain: SH2 homology <SH2>  
F:286-546/Domain: protein kinase homology <KIN>  
F:294-302/Region: protein kinase ATP-binding motif

Alignment Scores:  
Pred. No.: 5.4 Length: 1182  
Score: 92.00 Matches: 66  
Percent Similarity: 36.65% Conservative: 37  
Best Local Similarity: 23.49% Mismatches: 92  
Query Match: 6.67% Indels: 86  
DB: 2 Gaps: 14

US-09-658-824-808 (1-781) x A35962 (1-1182)

QY 36 AGGTCGTGATCTTCTCCGCTACTGAGACACGGCGGTAGTCCACAGCAGATCCAAAC 95  
::: ||||| |||||::: ||||| |||||  
Db 649 LysGlyGlyPhe-PheSerSerPheMetLysLysArgAsnAlaPro---ThrProProLy 667  
QY 96 TGGGAGTTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 134  
::: ||||| |||||::: ||||| |||||  
Db 667 sArgSerSerPheArgGluMetGluAsnGlnProHisLysLysTyrGluLeuThrGl 687  
QY 135 AGGCTTCGGAGGTTGTG-----TGGTCAGTGACTCAGAGTGA 173  
||| |||  
Db 687 yAsnPheSerValAlaSerLeuGlnHisAlaAspGlyPheSerPheThrProAlaGl 707  
QY 174 GAAGCCCTCGAAGTCGTGCTCTCTCAGCGGTGCGGCGCAGCCCATCGCTCTCTCTCT 233  
::: ||||| |||||::: ||||| |||||  
Db 707 nGlnGluAlaAsnLeuValPro-----LysCys-----TyrGlyGlySerPheAl 723  
QY 234 CGTCAGGCCATACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCT 293  
::: ||||| |||||  
Db 723 aGlnArgAsnLeuCysAsnAspGlyGly-----GlyGlyGlyGlySerGlyThrAl 741  
QY 294 GGGTGTGTGTG-----GGGTAT----- 312  
||| |||  
Db 741 aGlyGlyGlyTrpSerGlyThrGlyPhePheThrProArgLeuLeuLysLysThrLe 761  
QY 313 -----CCGAGTCCCGAAG 326  
||| |||||  
Db 761 uGlyLeuArgAlaGlyLysProThrAlaSerAspAspThrSerLysProPheProArgse 781  
QY 327 CACCTGGAACCCCGACAGAAGATTCTGGACTCCCGACACGGGACCAG----- 373

Db 781 rAsn-SerThrSerSerMetSerSerGlyLeuProGluGlnAspArgMetAlaMetThrI 801  
||| ::: ||||| |||||::: ||||| |||||  
QY 374 -----GAGAGGACCGCATGAGCGACACAC 398  
||| |||  
Db 801 euProArgAsnCysGlnArgSerLysLeuGlnLeuGluArg-Thr-----ValSer 817  
QY 399 ACAACACAGAACACACAGCAGCGATCCAGGAGCCAGTAATGGAGAGCCCAAAAAA--- 455  
|||::: |||||::: ||||| |||||  
Db 818 ThrSerSerGlnProGluGlnValAsnValAspArgAlaAsnAspMetLeuProLysLysSer 837  
QY 456 GAAGAACACGAGCAGTGAAGTCGGGATCCTACACCTGGCAGCAGACAGAGAAATCAG 515  
||| ||||| |||||::: |||||::: |||||  
Db 838 GluGluSerAlaAlaProSerArgGluArgProLysAlaLysLeuLeu-----ProArg 855  
QY 516 GATACAGCTGATGCCAGTCGGGACATGGAAGGTATCTGCAAGAGCTGCATCAGTCA 575  
::: |||||::: ||||| |||||  
Db 856 GlyAlaThrAlaLeuProLeuArgThrProSerGlyAspLeuAlaIleThrGluLys-As 875  
QY 576 AACACCGGGGATAATCTGGATTTCGGGTCGCGCTCAAGGTGAAGATAATACCTAAAGA 635  
|||::: |||||  
Db 875 pProGlyValGlyVal-----AlaGlyValAlaAlaAlaProLysGlyLysGl 892  
QY 636 GGAACAC-----TGTAAATGCCAGAAGCAGCTGAAGAG 669  
||::: |||||  
Db 892 uLysAsnGlyGlyAlaArgLeuGlyMetAlaGlyValProGluAspGlyGluGln 910

RESULT 34  
A35938  
profilaggrin - human (fragments)  
C:Species: Homo sapiens (man)  
C>Date: 14-Dec-1990 #sequence\_revision 02-Jul-1996 #text\_change 29-Sep-1999  
C:Accession: A35938  
R:Gan, S.Q.; McBride, O.W.; Idler, W.W.; Markova, N.; Steinert, P.M.  
Biochemistry 29, 9432-9440, 1990  
A:Title: Organization, structure, and polymorphisms of the human profilaggrin gene.  
A:Reference number: A35938; MUID:91064347  
A:Accession: A35938  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-2248 <GAN>  
A:CROSS-references: GB:J02929  
C:Genetics:  
A:Gene: GDB:FLG  
A:CROSS-references: GDB:119912; OMIM:135940  
A:Map position: 1q21-lq21  
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology  
F:246-569/Region: EF hand; epidermis; polymorphism; tandem repeat  
F:570-893/Region: filaggrin repeat  
F:1074-1397/Region: filaggrin repeat  
F:1573-1896/Region: filaggrin repeat

Alignment Scores:  
Pred. No.: 6.7 Length: 2248  
Score: 92.00 Matches: 52  
Percent Similarity: 30.22% Conservative: 16  
Best Local Similarity: 23.11% Mismatches: 79  
Query Match: 6.67% Indels: 78  
DB: 2 Gaps: 10

US-09-658-824-808 (1-781) x A35938 (1-2248)

QY 36 AGGTCGTGATCTTCTCCGCTACTGAGACACGGCGGTAGTCCACAGCAGATCCAAAC 95  
||| ||||| |||||  
Db 1394 ArgSerGlySerPheLeuTyrGlnValSerThrHisGluGlnSerGluSerHisGly 1413  
QY 96 TGGGAGTTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 155  
||| |||||  
Db 1414 TrpSer-----GlyThrSer----- 1418  
QY 156 GTCAGTGACTCAGATGAGAGGCCCTCGAAGTCGTGCTCCCTCTCATCGGTGCCACGCG 215  
||| |||||

```

Db 1419 -----ThrArgGlyArgGlnGlySerHisHisGluGlnAla 1430
QY 216 CCATGGACCTTCCTGCTCTCTG-----CAGGCG 242
Db 1431 GlnAspSer-----SerArgHisSerAlaSerGlnGluGlyGlnAspThrIleHisGly 1448
QY 243 CATNACTAGGAGAGGAGGCGGAGGAGTGGAGGGCTCAGCGGAGCTGGGGTGCTG 302
Db 1449 HisPro---GlyGlyHisSerAlaAspSerSerArgGlnSerGly-----1462
QY 303 TTGGGGGTATCCGAGTCCCAAGACCTCGGAACCCCGACAGAAATTCCTGGACTCCCA 362
Db 1463 -----ThrArgHisThrGlu-----1467
QY 363 GACGGGACAGGAGGAGCGGATGAGCGCACACACACAAACACAGAACCA-----413
Db 1468 ---SerSerSerArgGlyGlnAlaAlaSerHisGlnGlnAlaArgSerSerAlaGly 1486
QY 414 -----CACAGCCAGTCCAGGAGCCCAAGT-----GGA 443
Db 1487 GluArgHisGlySerHisHisGlnGlnSerAlaAspSerSerArgHisAlaGlyIleGly 1506
QY 444 GAGCCCCAAAGAAACAGCAGCTGAAAGT-----CGGGAT 482
Db 1507 HisGlyGlnAlaSerSerAlaValArgAspSerGlyHisArgGlyTyArgGlySerGln 1526
QY 483 CCTACACCTGGGAGCAGACAGAAAGATCAGGATACAGCTCAGATCCAGTCGCGGAC 542
Db 1527 AlaThrAspSerGluGlyHisSerGluAspSerAspThrGlnSerValSerAlaGlnGly 1546
QY 543 ATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATAAATCTGGATTGGG 602
Db 1547 GlnAlaGlyProHisGlnGlnAlaHisGlnLysSerAlaArgGlyGlnSerGlyGluSer 1566
QY 603 TTCGGGCGTCAAGGT 617
Db 1567 SerGlyArgSerGly 1571

RESULT 35
B87676
hypothetical protein CC3444 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: B87676
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87676
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-268 <STO>
A:Cross-references: GB:AE005673; NID:g13425162; PIDN:AAK25406.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3444

Alignment Scores:
Pred. No.: 6.36 Length: 268
Score: 91.50 Matches: 42
Percent Similarity: 34.81% Conservative: 13
Best Local Similarity: 26.58% Mismatches: 48
Query Match: 6.63% Indels: 56
DB: 2 Gaps: 9

US-09-658-824-808 (1-781) x B87676 (1-268)
QY 262 GGSCCGAGGAGTGGGGGCTCAGGGAAGCTGGGTGCTGTGGGGTAT-----312
Db 12 GlyPro---SerValGlyPheAsnAlaValLeuSerSerCysTrpGlyTyrlLeuAlaTyr 30
QY 313 ---CCGAGTCCAGAAAGACCTCGGAACCCCGACAGA-----AGATTCTGGACTCCC 360

```

```

Db 31 AlaProCysAlaArg-----GlnProAspArgThrArgProArgAspTrp-----45
QY 361 CACAGCGGACGAGGAGGAGGCGCATGAGCAGACACACAAACACACAGCCACACAGCC 420
Db 46 -----GlyArgGlyArgPheHisArgArgSerGlyThrAsnArgSer 61
QY 421 AGTCCCAGGAGCCCAATGAGAGCCCA-----AAAGAAGAACACAGCAGCTGAAAGT 476
Db 62 GlyProAlaAlaProProAlaGlnAlaProAlaAlaProAlaAlaGlnAla 81
QY 477 CGGGAT-----CCTACACCTGGCAGCAGACAGAAAGATCAG 515
Db 82 AlaProAlaAspGlnAlaAlaProValAlaAlaProAlaAlaGluThrAlaProAlaAla 101
QY 516 GATACAGCTGAGATCCAGTCGCGACATGGAA-----548
Db 102 AspThrTyrlValAlaProValArgGlyProArgThrThrAspProPheThrValArgLeu 121
QY 549 -----GGTGATCTGCAAGAGCTG 566
Db 122 LeuAspValLeuAspLysValCysLysProGlnValAlaGlyAspPheAlaGlnLeu 141
QY 567 CATCAGTCAACACCGGGGATAAATCTGGATTGGGTTCGGCGTCAAGGTGAA 620
Db 142 ValLysSer-----TyrGlyPheLysLysArgGlu 152

RESULT 36
S24169
mucin - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 10-Nov-1995
C:Accession: S24169
R:Huan, L.J.; Xu, G.; Forstner, G.; Forstner, J.
Biochim. Biophys. Acta 1132, 79-82, 1992
A:Title: A serine, threonine and proline-rich region near the carboxyl-terminus of a
A:Reference number: S24169; MUID:92379096
A:Accession: S24169
A:Molecule type: mRNA
A:Residues: 1-292 <HUA>

Alignment Scores:
Pred. No.: 6.4 Length: 292
Score: 91.50 Matches: 39
Percent Similarity: 40.16% Conservative: 10
Best Local Similarity: 31.97% Mismatches: 56
Query Match: 6.42% Indels: 17
DB: 2 Gaps: 5

US-09-658-824-808 (1-781) x S24169 (1-292)
QY 357 AGTCCAGAAATCTCTCGGGGTTCAGGTGCTTCTGGACTCGGATA---CCCCAACA 301
Db 90 ThrProThrSerThrSerThrValProThrThrGlySerThrSerSerLysProProThr 109
QY 300 GCACCCCGCTCGCCCTG---AGCCCTCCACTCTCGGCCCTCTCTCTCCCTAGTTAT 244
Db 110 GlySerSerThrProIleThrSerProSerThrPro-SerProProThrGlu-----126
QY 243 GGCCTGACGAGACAAAGAGTCCATGGCGTGGCACCGCATGAGAGGACGAGCTTC 184
Db 127 -SerThrThrLeuSerSerThrPro-----ValThrThrTh 138
QY 183 GAGGGCTTCTCACTCTGAGTCACTGACCACACAAACCCCGGAAGCCTGCTGGTTCCTC 124
Db 138 rAlaThrSerSerThrThrSerSerProGlyThrThr-----SerProPheAlaThrSe 156
QY 123 TTCCTCTCACTCACACTTCAACTCCCACTGGATCGCTGTGGACCTACCCGCGTGT 64
Db 156 rSerValSerSerThrProProSerProSerProSerSerGlyProThrThrSerSe 176
QY 63 CTCAGTAGCGGAGAAAGATCCAGACCTCAGGGACCCGAGTCGCCGGCTCACAGCTCCGC 4

```

Db 176 rGlyMetProThrSerSerLysThrThrThrGlyProThrSerProThr-ThrArgProp 196  
QY 3 CG 2  
Db 196 ro 196  
RESULT 37  
S39406  
homeotic protein otx1 - human  
C:Species: Homo sapiens (man)  
C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 15-Oct-1999  
C:Accession: S39406  
R:Simone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; E  
EMBO J. 12, 2735-2747, 1993  
A:Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid  
A:Reference number: S35345; MUID:93327763  
A:Accession: S39406  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-354 <Sim>  
C:Genetics:  
A:Gene: otx1  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:39-95/Domain: homeobox homology <HOX>  
Alignment Scores:  
Pred. No.: 6.49 Length: 354  
Score: 91.50 Matches: 50  
Percent Similarity: 36.46% Conservative: 20  
Best Local Similarity: 26.04% Mismatches: 59  
Query Match: 6.42% Indels: 63  
DB: 2 Gaps: 10  
US-09-658-824-808 (1-781) x S39406 (1-354)  
QY 622 TCTTCACCTTGACGCGGAAACCCAAATCCAGATTATCCCGGTTTGTACTGATGCAGC 563  
Db 112 SerSerProValArgGluSerSerGlySerGluSerSerGlyGlnPhe----- 127  
QY 562 TCTTCGATCACCCTCCATCGCGCACTGGGACTGCTGAGTCTCAGTGTATCCTGATCTCTCT 503  
Db 128 -----ThrProAlaValSerSerSerAlaSerSerSerSerAlaSerSer 145  
QY 502 GTCTGCTGCCAGGTGTAGGATCCGACTTTCAGCTGTGTTCTTCTTTTGGGGTCT 443  
Db 146 SerAlaAsnProAla-----AlaAlaAlaAlaAlaGlyLeuGlyGlyAsn 160  
QY 442 CCATTACTGGGCTCTGGGACTGGCTGTGCTGTCTGTGTTGTGTGCTGCTCATGCC 383  
Db 161 Pro-----ValAlaAlaAla 165  
QY 382 GTCCCTCTCTGTGTCCTGTGGGAGTCCAGATCTTCTGTGGGGTTCCAGTGTCTC 323  
Db 166 SerSerLeuSerThrProAlaAlaSerSer-----Ile 176  
QY 322 TGGGAC-----TCGGATACCCCAACAGCACCCAGCTTCG-----CGTAGGCC 278  
Db 177 TrpSerProAlaSerIleSerProGlySerAlaProAlaSerValSerValProGluPro 196  
QY 277 CTC-----CACTCTCTCGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 230  
Db 197 LeuAlaAlaProSerAsnThrSerCysMet-----GlnArg 208  
QY 229 AAGAAGGTCCATGGCGGTGGCCAGCCGATGAGAGGACGACGACTT-----CGAGGG 179  
Db 209 SerValAlaAlaGlyAlaAlaThrAlaAlaAlaSerTyrProMetSerTyrGlyGlnGly 228  
QY 178 CTTTCTCACTGTGACTGACCTGACCAACACACCTCC----- 143  
Db 229 GlySerTyrGlnGlyTyrProThrProSerSerSerTyrPheGlyGlyValAspCys 248  
QY 142 GGAAGCTGCTGTTCCCTTCTCACTCTCACTCTCACTCACAC 107

Db 249 SerSerTyrLeuAlaProMetHisSerHisHis 260  
RESULT 38  
A45988  
dentin matrix acidic phosphoprotein AG1 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 25-Oct-1996  
C:Accession: A45988  
R:George, A.; Sabay, B.; Simonian, P.A.; Veis, A.  
J. Biol. Chem. 268, 12624-12630, 1993  
A:Title: Characterization of a novel dentin matrix acidic phosphoprotein. Implication  
A:Reference number: A45988; MUID:93286101  
A:Accession: A45988  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-489 <Geo>  
A:Experimental source: odontoblast-pulp fibroblast, Sprague-Dawley  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:133593, NCBI:P:133595)  
C:Keywords: phosphoprotein  
Alignment Scores:  
Pred. No.: 6.64 Length: 489  
Score: 91.50 Matches: 60  
Percent Similarity: 35.24% Conservative: 20  
Best Local Similarity: 26.43% Mismatches: 94  
Query Match: 6.63% Indels: 53  
DB: 2 Gaps: 10  
US-09-658-824-808 (1-781) x A45988 (1-489)  
QY 22 GACTCGGGTCCCTGAGGTCTGGGTCTTCTCCGCTACTGAGACACGGCGGTAGGTCCA 81  
Db 89 AspSerGly-----AspAspThrPheGlyAspGluAspAsnGlyProGlyPro 104  
QY 82 CAGCAGATCCAACTGGGAGTTGAAGTGTGAGTGAGAGTGAAGAGAACAGCAGCGTTC 141  
Db 105 GluGluArgGlnTrpGlyGlyProSerArgLeuAspSerAspGluAspSerAla----- 122  
QY 142 CGAGGGTTGTGGTGTGAGTCTGAGTCTGAGTGAAGAGCCCTCGAGTGTGCTGCTCTC 201  
Db 123 -----AspThrThrGlnSerSerGlu-AspSerThrSerGlnGluAsnSe 137  
QY 202 ATCGGTGTCACGCCCATGGACCTTCTT-----GTCTCGTCACGCC 243  
Db 137 rAlaGlnAspThrProSerAspSerLysAspHisSerAspGluAlaAspSerArgPr 157  
QY 244 ATACTAGG-----AGGAAGAGGCGCGAGGAGTGGAGG---GGCTCAGGCG 288  
Db 157 oGluAlaGlyAspSerThrGlnAspSerGlu-SerGluGluTyrArgValGlyGlyGly 177  
QY 289 AAGCTGGGTGCTGTTGGGGTATCCGAGTCCAGAGACACCTGGAAACCCCGACAGAGA 348  
Db 177 erGluGlyGluSerSerHisGlyAspGlySerGluPheAspAspGluGlyMetGlnSer- 196  
QY 349 TTCTGGACTCCCGACAGCGGACCCAGG-----AGAGGACGGCATGAGCCACACACAA 402  
Db 197 -----AspAspProGlySerThrArgSerAspArgGly-----HisThrA 210  
QY 403 ACAGAAACACACAGCAGCTCCCGAGGAGCCAGTATGAGAGCCCAAAAGAGAGAAC 462  
Db 210 rgMetSerSerAlaAspIleSerSerGluGluSerLysGlyAsp----- 224  
QY 463 CAGCAGCTGAAGTTCGGGATCCTACACCTGGGAGCAGACAGAGAGAGATCAGGATACA- 521  
Db 225 -----HisGluProThrSerThrGlnAspSerAspSerGluAspValG 240  
QY 522 -----CCTGAGTATCCAGTCCCGCAGCATGGAAGGTG 552  
Db 240 luPheSerSerArgLysSerPheArgArgSerArgValSerGluGluAspAspArgGly 260  
QY 553 ATCTGCAAGAGCTGCATCAGTCAACACCGGGGATAAATCTGGATTGTTGGGTCGCGGTC 612

Db 260 luLeuAlaAspSerAsnSerArgGluThrGlnSerValSerThrGluAspPheArgSerL 280  
QY 613 AAGGTGAAGATAATACC 629  
Db 280 ysGluGluSerArgSer 285  
RESULT 39  
T31422  
C-terminal domain-binding protein RA9 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 29-Oct-1999 #sequence\_revision 29  
A:Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts with Ruvruev, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; Cordeiro, Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996  
A:Reference number: 221024; MUID:96293459  
A:Accession: T31422  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1473 <YUR>  
A:Cross-references: EMBL:U49057; NID:g1438533; PID:g1438534; PIDN:AAC52658.1  
A:Experimental source: hippocampus

Alignment Scores:  
Pred. No.: 7.94 Length: 1473  
Score: 91.00 Matches: 46  
Percent Similarity: 37.50% Conservative: 17  
Best Local Similarity: 27.38% Mismatches: 46  
Query Match: 6.59% Indels: 59  
DB: 2 Gaps: 9  
US-09-658-824-808 (1-781) x T31422 (1-1473)  
QY 128 AACGAGCAGGCTCCGAGGGTTGTGTCAGTACTCAGAGTGAAGCCCTCGAAG 187  
Db 359 SerGluSerValSerGlyGlyLeuIleGlyAsp 372  
QY 188 TCGTCGCTCCCTCATCGGTP 208  
Db 373 AsnSerGlyLeuSerCysGlyAspArgThrAlaLeuArgCysLeuProAlaGlnIleVal 392  
QY 209 GCCAGCCCATG-----GACCTTCTGTCGTCAGCGCCATTAACTAGGAGGAGGAG 262  
Db 393 GlnThrProValArgSerAspSerValSerProArgSerGlyLeu-SerGlyAsnLe 412  
QY 263 GCCGAGGAGTGGAGGGCTCAGCGAAGTGGGTGCTGTGGGGTATCCGAGTCCA 322  
Db 412 userAspGluSerArg----- 417  
QY 323 GAAGCA:CTGGAACCCGACAGAAGATTCTGGACTCCCA-----GACGGGAC 370  
Db 418 -----ProLysTrpLysHisSerAsnSerProArgLeuAsnGlySerAsnVa 433  
QY 371 CAGGAGGAGCGGATGAGGACACACACAAACACAGACAGCCAGTCCAGGA 430  
Db 433 lArgValGlySerAla---SerTrpLysThrMetThrHisSerAsnPheProSerLysAs 452  
QY 431 GCCCAGTAATGAGAGCCCAAAAGAGAACACAGCAGCTGAAAGTCGGGATCCTACACC 490  
Db 452 nIleAlaProGlyHisProGlnLysThrAspProArgProAspPheSerLysLeuPr 472  
QY 491 T-----GGCAGCAGACAGAAAGATCAG----- 515  
Db 472 oArgileProLysIleArgAspGlySerAsnSerThrGlnAspGlnAlaProAlase 492  
QY 516 ----GATACAGTGCAGTCCCA 533  
Db 492 rGlyGlnThrValGluLeuPro 499

RESULT 40  
S70009  
glutamate/proline-rich protein (clone 53.1.1.1) - rat

C:Species: Rattus norvegicus (Norway rat)  
C:Date: 06-Dec-1996 #sequence\_revision 13  
A:Title: Cloning and characterization of cDNAs for novel proteins with glutamic acid  
A:Reference number: S70009; MUID:96221285  
A:Accession: S70009  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-221 <GEE>  
A:Cross-references: EMBL:U40627; NID:g1184693; PIDN:AAB05667.1; PID:g1184694  
Alignment Scores:  
Pred. No.: 7.67 Length: 221  
Score: 90.50 Matches: 36  
Percent Similarity: 37.32% Conservative: 17  
Best Local Similarity: 25.35% Mismatches: 47  
Query Match: 6.56% Indels: 42  
DB: 2 Gaps: 6  
US-09-658-824-808 (1-781) x S70009 (1-221)  
QY 293 TGGGGTCTCTGTGG-----GGTATC 313  
Db 93 TrpGlnHisValGlyProGlyTyrArgAspArgSerTyrAspProCysProGlyHis 112  
QY 314 CGAGTCCAGAACACCTGGAAACCCGACAGAGATTCTGGACTCCCGACGGGACAG 373  
Db 113 TrpThrProGluAlaProSerSerGlyThrThrCysProGlyLeuProArgAlaSerGlu 132  
QY 374 GAGAGGAC-----GGCATGAGCGACACACAAACACAGACACAGCCAGTCCCA 427  
Db 133 GluGluLileGlyGlyProGluAspSerGluAla-ValGlnProArgThrProGluG 152  
QY 428 G-----GAGCCAGTATGGA-----GA 445  
Db 152 uProGluLeuGluAlaGluAlaThrLysGlyAspGluProAspLeuGluGlnGluMetG 172  
QY 446 GCCCAAAAGAAAGAACCCAGCAGCTGAAAGTCGGGATCCTACACTGGGACAGACAGA 505  
Db 172 uProGluProGluProGluValGluProGluProGluProGluProGluProGluProG 192  
QY 506 AGAAGATCAGATACAGTCCAGATCCAGTCCGCGACATGGAAGGTGATCTGCAAGAGCT 565  
Db 192 uProGluProGluProGluProGluProGluProGluProGluProGluProGluProG 211  
QY 566 GCATCAGTCAACACACCCGGGATAAATCTGGATTTCGGGTCGCGCTCAAGGTGAAGATA 625  
Db 212 -----GlyAspGluSer-----GluGlyCysGluAs 220  
QY 626 TACC 629  
Db 220 nThr 221

Search completed: October 18, 2002, 09:45:49  
Job time : 58 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	550	39.9	146	1	GGD2_HUMAN	Q9hd64	homo sapien
2	138.5	10.0	102	1	GGC1_HUMAN	O60829	homo sapien
3	114	8.3	908	1	SRCA_RABIT	P13666	oryzctolaqus
4	110	8.0	146	1	GBL1_HUMAN	O75459	homo sapien
5	106	7.7	246	1	GRPA_RAT	P08568	rattus norv
6	106	7.7	527	1	GI9P_HUMAN	P14314	homo sapien
7	103.5	7.5	2472	1	NCR2_MOUSE	Q9wu42	mus musculus
8	103	7.5	247	1	GRPB_RAT	P08462	rattus norv
9	103	7.2	363	1	CYB_TRYBB	P00164	trypanosoma
c	99.5	7.2	503	1	DMP1_MOUSE	O55188	mus musculus
c	110	99.5	7.0	598	NR41_HUMAN	P22736	homo sapien
12	98	7.1	1736	1	CA2B_HUMAN	P13942	homo sapien
c	13	97	371	1	CYB_LEIFA	P14548	leishmania
14	96.5	7.0	0	277	HMPH_CHICK	Q05502	gallus gall
c	15	96	6.7	348	YF35_STRFR	P20186	streptomyces
16	95.5	6.9	117	1	GGE4_HUMAN	Q13068	homo sapien
17	95.5	6.9	117	1	GGE5_HUMAN	Q13069	homo sapien
18	95.5	6.9	117	1	GGE6_HUMAN	Q13070	homo sapien

```

Best Local Similarity: 52.19%      Mismatches: 1
Query Match: 39.86%      Indels: 108
DB: 1      Gaps: 1

US-09-658-824-808 (1-781) x GGD2_HUMAN (1-146)

QY 2 CGGCGGAGCTGTGAGCGCGGCGACTCGGTCCTGAGCTGAGTCTTTCCTCGCTACTG 61
   |||||||
DB 27 ATGAGS2rCysGluProAlaThrArgValProGluValrPilleuSerProLeuLeu 46
   |||||||
QY 62 AGACAGCGGGGTAGTCCACAGGCGAGATCCAACTGGGAGTTGAAGTGTGAGTGAGAGTG 121
   |||||||
DB 47 ArgHisGly----- 49

QY 122 AAGAGGAACACAGAGCTTCGCGAGGGTGTGTGTCAGTCACTCAGAGTGAGAGAGGCC 181
   |||||||
DB 49 ----- 49

QY 182 TCGAGTCTGCTCCTCTCTATCGCGGTGCCGCCCTTGACCTTCTTGTCTGTCTGTCACGG 241
   |||||||
DB 49 ----- 49

QY 242 CCATAACTAGGAGGAGGAGGCGGAGGTGGAGGGCTCAGCGGAAGCTGGGGTGTCT 301
   |||||||
DB 49 ----- 49

QY 302 GTTGGGGGTATCCGAGTCCAGAGCACTTGGAAACCCGACAGAAAGATTCTGACTCCCC 361
   |||||||
DB 49 ----- 49

QY 362 AGACGGGACGAGGAGGCGCATGAGCGGACACACAAACACAGAACACACAGCCA 421
   |||||||
DB 50 -----Gly--HisThrGlnGlnAsnHisThrAlaAs 60

QY 422 GTCCAGGAGCCCAATGAGAGCCCAAAAGAAAGAACAGCAGCTGAAAGTCGGGA 481
   |||||||
DB 60 erProArgSerProValMetGluSerProLysLysLysAsnGlnGlnLeuLysValGlyI 80

QY 482 TCCTACACTGGGACGACAGAGAGATCAGATACAGTCACTGAGATCCAGTCGCGGA 541
   |||||||
DB 80 leLeuHisLeuGlySerArgGlnLysLysIleArgIleGlnLeuArgSerGlnCysAla 100

QY 542 CATGAAGGTGATCTGCAAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGATTGG 601
   |||||||
DB 100 hrTriLysValIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeu 120

QY 602 GTTCCGGCTCAAGTGAAGATATACCTAAGAGGAACACTGTAAATGCCAAGACAG 661
   |||||||
DB 120 lyserGlyValLysValLysIleIleProLysGluGluHisCysLysMetProGluAla 140

QY 662 GTGAAGAGCAACCAAGTT 681
   |||||||
DB 140 lyGluGluGlnProGlnVal 146

RESULT 2
GGC1_HUMAN
ID GGC1_HUMAN STANDARD; PRT; 102 AA.
AC 060829;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE G antigen family C1 protein (prostate-associated gene protein 4)
DE (PAGE-4) (PAGE-1) (JM27) (GAGE-9).
GN GAGE1 OR PAGE4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98393718; PubMed=9724777;
RA Brinkmann U., Vasmatzis G., Lee B., Yerushalmi N., Essand M.,
RA Pastan I.;

```

```

RT *PAGE-1, an X chromosome-linked GAGE-like gene that is expressed in
RL normal and neoplastic prostate, testis, and uterus.;
RN Proc. Natl. Acad. Sci. U.S.A. 95:10757-10762(1998).
RP [2]
RP SEQUENCE FROM N.A.
RA Strom T.M., Nyakatura G., Hellebrand H., Drescher B., Rosenthal A.,
RL Meindl A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: UNKNOWN.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN NORMAL MALE AND
CC FEMALE REPRODUCTIVE TISSUES, PROSTATE, TESTIS, FALLOPIAN TUBE,
CC UTERUS, AND PLACENTA, AS WELL AS IN PROSTATE CANCER, TESTICULAR
CC CANCER, AND UTERINE CANCER.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF275258; AAF88037.1; -
DR EMBL; AJ005894; CAA06751.1; -
DR EMBL; AF238380; AAF62541.1; -
DR MIN; 300287; -
SQ SEQUENCE 102 AA; 11153 MW; CE5D07AFBF73301B CRC64;

Alignment Scores:
Pred. No.: 0.00201      Length: 102
Score: 138.50      Matches: 34
Percent Similarity: 55.70%      Conservative: 10
Best Local Similarity: 43.04%      Mismatches: 31
Query Match: 10.04%      Indels: 4
DB: 1      Gaps: 1

US-09-658-824-808 (1-781) x GGC1_HUMAN (1-102)

QY 441 GGAGAGCCCCAAAAGAACAGACAGCAGCTGAAAGTCGGGATCTACACCTGGCGCAG 500
   |||||||
DB 28 GlyGluSerGlnGlnGluGluProThrAspAsnGlnAspIleGluProGlyGlnGlu 47

QY 501 ACAGAGAAGATCAGGATACAGCTGAGATCCAGTCGCGCAGCATGGAAGGTGATCGCA 560
   |||||||
DB 48 ArgGlu-----GlyThrProIleGluGluArgLysValGluGlyAspCysGln 64

QY 561 GAGCTGCATCAGTCAACACCGGGGATAAATCTCGATTGGTTCCGCGCTCAAGGTGAA 620
   |||||||
DB 65 GluMetAspLeuGluLysThrArgSerGluArgGlyAsp-GlySerAspValLysGluLy 84

QY 621 GATAATACCTAAAGAGGAACACACTGTAAATGCCAGAACAGGTCGAGCAACCA 675
   |||||||
DB 84 sThrProAsnProLysHisAlaLysThrLysGluAlaGlyAspGlyGlnPro 102

RESULT 3
SRCA_RABIT
ID SRCA_RABIT STANDARD; PRT; 908 AA.
AC P13666;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sarcalumenin precursor.
DE Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=89345602; PubMed=2762314;
RA Leberer E., Charuk J.H.M., Green N.M., MacLennan D.H.;
RT *Molecular cloning and expression of cDNA encoding a luminal calcium
RT binding glv opote'n from sarcoplasmic reticulum.;

```

QY	340	GACAGAAAGATTCTGCTGCTCC	CC	CAAGAGGGGACACAGAGGAGCGCATGAGCGACACACA	399
		:::     :::     :::			
Db	394	rSerGluGluAlaGlyAlaAlaSerGluGluAlaSer-GlyThrAlaGlyLeuGlyGluG	414		
QY	400	CAACACACAGAACCCACAC	-----	AGCCAGTCCCAGG	429
				:::	
Db	414	luGluThrGlnProSerThrGluGlyLeuAspSerGlyProAlaGlySerGlnAlaGlnA	434		
QY	430	AGCCCACTAATGGAGAGCCGCCAAAAGAAAGAACACGACGCTGAAGCTCGGATCCTACAC	489		
		:::     :::			
Db	434	spThrGluAlaGluGluProGluGluHisGlnGlyProGluSerProIleThrAlaP	454		
QY	490	CTGGGCGACGACACAGAAGAATCATCAGGTACAGCTGCAGTCCCGAGTCGCGACATGGAAG	549		
Db	454	roGlnGluThrGluAspValSerGlu	-----	GluValProMetArgasp	469
QY	550	GTGATCTGCAAGAGCTGCATCATAGTCAACACACGGGGATAATCTGGATTTGGTTCCGGC	609		
Db	469	-----			469
QY	610	GTCAAGGTGAAGATATACCTTAAGAGGACACTGTAATAATGCCAAGACGAGGTGAAGAG	669		
Db	470	-----		ArgSerHisIleGlnL	475
QY	670	CAACCACAAAGTTTAAATGAAGACAAG	695		
Db	475	ysThrLeuMetLeuasnGluAspLys	483		
RESULT 4					
GGBI_HUMAN					
ID	GGBI_HUMAN		STANDARD;	PRT;	146 AA.
AC	O75459;				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	K antigen family B 1 protein (prostate-associated gene protein 1)				
DE	(PAGE-1) (GAGE-9) (AL5).				
GN	GAGEB1 C <sub>H</sub> PAGE1 OR GAGE9.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI_Taxid:9606;				
RP	[1]				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE-98316329; PubMed-9651357;				
RA	Chen M.E., Lin S.-H., Chung L.W.K., Sikes R.A.;				
RT	*Isolation and characterization of PAGE-1 and GAGE-7: new genes				
RT	expressed in the LNCA prostate cancer progression model that share				
RT	homology with melanoma-associated antigens.;				
RL	J. Biol. Chem. 273:17618-17625(1998).				
CC	-1- FUNCTION: UNKNOWN.				
CC	-1- TISSUE SPECIFICITY: ISOLATED FROM PROSTATE CANCER CELL LINES;				
CC	EXPRESSION ASSOCIATED WITH PROGRESSION TO ANDROGEN INSENSITIVE				
CC	PHENOTYPE. EXPRESSED IN NORMAL TESTIS AND AT LOWER LEVEL IN NORMAL				
CC	PLACENTA.				
CC	-1- SIMILARITY: BELONGS TO THE GAGE FAMILY.				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collabora				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstati				
CC	the European Bioinformatics Institute. There are no restrictions on				
CC	use by non-profit institutions as long as its content is in no				
CC	modified and this statement is not removed. Usage by and for commerc				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/annou">http://www.isb-sib.ch/annou</a>				









CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
 CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIIB.  
 CC -|- SUBCELLULAR LOCATION: Nuclear  
 CC -|- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -|- TISSUE SPECIFICITY: UBQUITOUS. ALSO WIDELY EXPRESSED IN EARLY  
 CC EMBRYOS.

CC -|- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
 CC DOMAINS (ID1 AND ID2).

CC -|- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND  
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES  
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
 CC SPECIFICITY.

CC -|- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.

CC -|- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.

CC -|- SIMILARITY: CONTAINS 2 CORNR BOX.

CC -|- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF113001; AAD20944.1; -

CC EMBL: AF113002; AAD20945.1; -

CC EMBL: AF125671; AAD22972.1; -

CC MGD: MGI:1337080; Ncor2.

CC InterPro: IPR001005; Myb\_DNA\_Bind.

CC Pfam: PF00249; myb\_DNA-binding; 2.

CC SMART: SM00395; SANT; 2.

CC PROSITE: PS00090; MYB\_3; 1.

CC Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 CC Coiled coil; Alternative splicing.

CC COILED COIL (POTENTIAL).

CC INTERACTION WITH SIN3A/B (BY SIMILARITY).

CC SANT-A (POTENTIAL).

CC MYB.

CC COILED COIL (POTENTIAL).

CC COILED COIL (POTENTIAL).

CC PRO-RICH.

CC PRO-RICH.

CC CORNR BOX OF ID1.

CC CORNR BOX OF ID2.

CC POLY-GLN.

CC POLY-ALA.

CC POLY-PRO.

CC MISSING (IN ISOFORM BETA).

CC M -> RL (IN REF. 2).

CC PPMYDA -> RHVRR (IN REF. 2).

CC D -> H (IN REF. 1; AAD20944).

CC T -> M (IN REF. 1; AAD20944).

CC V -> A (IN REF. 2).

CC HHLPHRLTLRMNKKRLQLPRLPRRSRLPRRSWME  
 CC -> PSPAPPTATVDKDEAPAAPQTEDAKQKSEAEI  
 CC DVG (IN REF. 2).

CC E -> K (IN REF. 1; AAD20945).

CC E -> K (IN REF. 1; AAD20945).

CC E -> K (IN REF. 1; AAD20945).

CC E -> K (IN REF. 1; AAD20945).

CC S -> F (IN REF. 1; AAD20944).

CC I -> IQ (IN REF. 1; AAD20944).

CC PKLPTPPRWSSGLPFPI -> QSYRLSPHAGHRLPSSH  
 CC (IN REF. 2).

CC PHAADPSA -> TRADPL (IN REF. 2).

CC CONFLICT 1073 1080

CC CONFLICT 1073 1080

CC CONFLICT 1073 1080

CC CONFLICT 1073 1080

CC CONFLICT 1073 1080

CC CONFLICT 1073 1080

CC CONFLICT 1073 1080

CC CONFLICT 1073 1080

FT CONFLICT 1133 1133 MISSING (IN REF. 2).  
 FT CONFLICT 1149 1149 MISSING (IN REF. 2).  
 FT CONFLICT 1157 1157 G -> E (IN REF. 2).  
 FT CONFLICT 1172 1201 SPATSGITKGLPSTRAADGPSYRGSITHG -> APPVVEA  
 FT CONFLICT 1172 1201 GSATSGITKGLPSTRAADGPSYRGSITHG (IN REF. 2).  
 FT CONFLICT 1696 1696 A -> S (IN REF. 2).  
 FT CONFLICT 1855 1857 MISSING (IN REF. 2).  
 FT CONFLICT 1909 1909 A -> P (IN REF. 2).  
 FT CONFLICT 1913 1913 A -> G (IN REF. 2).  
 FT CONFLICT 1923 1923 G -> A (IN REF. 2).  
 FT CONFLICT 1956 1956 N -> S (IN REF. 2).  
 FT CONFLICT 1968 1968 A -> G (IN REF. 2).  
 FT CONFLICT 2195 2196 TA -> AV (IN REF. 2).  
 FT CONFLICT 2213 2214 LE -> SK (IN REF. 2).  
 FT CONFLICT 2224 2224 T -> A (IN REF. 2).  
 SQ SEQUENCE 2472 AA; 270856 MW; 2A58F4DFB79285B CRC64;  
 Alignment Scores:  
 Pred. No.: 1.34 Length: 2472  
 Score: 103.50 Matches: 67  
 Percent Similarity: 36.56% Conservatives: 35  
 Best Local Similarity: 24.01% Mismatches: 82  
 Query Match: 7.50% Indels: 96  
 Db: 1 Gaps: 13  
 US-09-658-824-808 (1-781) x NCR2\_MOUSE (1-2472)  
 QY 25 TCGGGTCCCTGAGGCTCGGATCTTCTCCGCTACTGAG----- 63  
 Db 734 SerGlyProAlaAlaValAsnSerAspThrGluSerValProSerProArgSer 753  
 QY 64 -----ACACGGCGGGTAGTCCACAGGCGAGATCAACTGGGAGTTGAAGTGTGAGTGAG 117  
 Db 754 GluAlaThrLysAspThrGlyProLys-----ProThrGlyThrGluAlaLeuProAla 771  
 QY 118 ACTGAAGAGGAACACAGCAGGCTT-CCGAGGAGTTGTGTGTCAGTCACTCAGAG----- 170  
 Db 772 AlaThrGlnProProValProProGluProGluProAlaValAlaProAlaGluProSer 791  
 QY 171 -----TGAGAAGGCCCTCGAAGTCGTCTGCTCCTCATCGTCCGTCGCGGACGCCCA 218  
 Db 792 ProValProAspAlaSerGlyProProSerProGluProSerHisLeuProHisPro 811  
 QY 219 -----TGGACCTCTTGTCTCTCAGGCCCAATACTAGGGAGGAAGAGGCGCCGAG 269  
 Db 812 ArgLeuLeuThrArgMetAsnLysLysProArg----- 823  
 QY 270 GAGTGGAGGGCTCAGGCGAAGCTGGGGTGTGTGGGGTATCCGAGTCCAG----- 323  
 Db 824 -----LeuLeuGlnLeuProArgGlnArgMetPro 833  
 QY 324 -----AAGCACCTGGAAACCCGACAGAGAGATTTCTGGACTCCCGACGGGAGGAGAGA 377  
 Db 834 ArgSerArgSerLeuArgProArgSerMetIrpGluLysProGlu----- 849  
 QY 378 GGGACGGCATGAGCGACACACAAACACACACACACACAGCCAGTCCAGAGGCC----- 434  
 Db 850 -----GluProGluAlaSerGluGluProGlu 859  
 QY 435 AGTAATGGAGAGCCCAAAAGAACACAGCAGCTGAAAGTGGGATCTTACACCTGGG 494  
 Db 860 SerValLysSerAspHisLysGluGluThrGluGluProGluAspLysAlaLysGly 879  
 QY 495 CACGACAGAGAAGATCAGGATACAGTGTGATCCAGTCCGCGCATGAA----- 548  
 Db 880 -----ThrGluAlaLeuGluThrValSerGluAlaProLeuLysValGluGluAlaGly 897  
 QY 549 -----GGTGTCTGCAAGAGCTGCATCAGTCA 575  
 Db 898 SerLysAlaAlaValThrLysLysSerSerGlyAlaThrGlnAspSerAspSerSer 917  
 QY 576 AACACC-----GGGGATAAATCTGGATTTGGG 602  
 Db 576 AACACC-----GGGGATAAATCTGGATTTGGG 602

Db 918 AlaThrCysSerAlaAspGluValAspGluProGluGlyGlyAspLys----- 933

QY 603 TTCGGCGTCAAGTGAAGATAATACCTAAAGAGGACACACTGTAATAATGCCAGAGCAGG 662

Db 934 -----GlyArgLeuLeuSerProArgProSerLeuLeuThrProAlaGlyAsp 949

QY 663 TGAAGACAACCACTCACT-----TTAAATCAAGACACAGCTGAACACACGC 707

Db 950 ProArgAlaSerThrProGlnLysProLeuAspLeuLysGlnLeuLysGlnArg 968

RESULT 8

GRPB\_RAT

ID GRPB\_RAT STANDARD; PRT; 247 AA.

AC P08462;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Submandibular gland secretory GLX-rich protein CB precursor (GRP-CB)

DE (Contiguous repeat polypeptide) (CRP).

GN GRP-CB.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=91139639; PubMed=3558393;

RA Heinrich G., Habener J.F.;

RT \*Genes encoding proteins with homologous contiguous repeat sequences

RT are highly expressed in the serous cells of the rat submandibular

RT gland.\*;

RL J. Biol. Chem. 262:5262-5270(1987).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar;

RX MEDLINE=91139639; PubMed=1995617;

RA Cooper L.F., Elia D.M., Tabak L.A.;

RT \*Secretagogue-coupled changes in the expression of glutamine/glutamic

RT acid-rich proteins (GRPs). Isoproterenol induces changes in GRP

RT transcript expression and changes in isoforms secreted.\*;

RL J. Biol. Chem. 266:3532-3539(1991).

CC -1- FUNCTION: GRP PROTEINS HAVE A MARKED AFFINITY FOR HYDROXYAPATITE.

CC THEY MAY PLAY A ROLE IN THE FORMATION OF THE PROTECTIVE ACQUIRED

CC PELLICLE AT THE SALIVA-TOOTH INTERFACE.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: SUBMANDIBULAR GLAND ACINAR CELLS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M31032; AAA40969.1; -

DR EMBL; M58654; AAA41279.1; ALT\_SEQ.

DR PIR; A29545; A29545.

DR PIR; B38647; B38647.

KW Submandibular gland; Saliva; Repeat; Signal; Multigene family.

FT SIGNAL 1 18

FT CHAIN 19 247

FT FT

FT DOMAIN 67 181

FT REPEAT 67 89

FT REPEAT 90 112

FT REPEAT 113 135

FT REPEAT 136 158

FT REPEAT 159 181

FT REPEAT 1 207

FT SIMILAR 247 AA; 26948 MW; 72A5075BCCC314E4 CRC64;

SQ SEQUENCE

Alignment Scores:

Pred. No.: 1.19 Length: 247

Score: 103.00 Matches: 39

Percent Similarity: 40.29% Conservative: 17

Best Local Similarity: 28.06% Mismatches: 50

Query Match: 7.46% Indels: 33

DB: 1 Caps: 6

US-09-658-824-808 (1-781) x GRPB\_RAT (1-247)

QY 306 GGGGTATCCGAGTCCCGAAGACACCTGTAACCGACAGAGATTTCTGGACTCCCGAC 365

Db 74 GlySerGluGluGluGlnGlnGlnGluProThrGlnAlaGluAsnGlnGluPro 93

QY 366 GGGACAGAGAGAGGCGCATGAGCGACACACACAAACACACACAGCCAGTC 425

Db 94 AlaThr-----SerGlySerGluGluGlnGlnGln 104

QY 426 CAGGAGCCCACT-----AATGGAGAGCCC----- 449

Db 105 GlnGluProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySerGluGlu 124

QY 450 ---CAAAAGAAGAACACGCA---GCTGAAAGTCGGATCTACACCTGGCGCAGCACA 503

Db 125 GlnGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySer 144

QY 504 GAAGAAGATCAGGATACAGCTGAGATCCAGTCGCGACATGGAAGTGATCTCAAGAG 563

Db 145 GluGluGluGlnGlnGlnGlu-----SerThrGlnAlaGluAsnGlnGlu 160

QY 564 CTGCATCAGTCMAACACCGGGGATAAATCTGGATTTGGGTTCCGGCTCAAGGTGAAG 623

Db 161 -----ProSerAspSerAlaGlyGlnGlnGlnGlnProGluGlu 175

QY 624 AATACCTAAAGAGGACACCTGTAATAATCCAGAGCAGGTGAAGACGACCAACAA 678

Db 176 GlyAsn-ValGluSerProProSerProGluAsnSerGlnGlnGlnProGln 193

RESULT 9

CYB\_TRYBB

ID CYB\_TRYBB STANDARD; PRT; 363 AA.

AC P00164;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Cytochrome b.

GN COB OR CYTB.

OS Trypanosoma brucei brucei.

OG Mitochondrion.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI\_TaxID=5702;

RN [1]

RP SEQUENCE FROM N.A. (KINETOPLAST).

RX MEDLINE=84041494; PubMed=6314266;

RA Benne R., de Vries B.F., van den Burg J., Klaver B.;

RT \*The nucleotide sequence of a segment of Trypanosoma brucei

RT apocytchrome b and some unusual unassigned reading frames.\*;

RL Nucleic Acids Res. 11:6925-6941(1983).

CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE

CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A

CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL

CC COUPLED TO ATP SYNTHESIS.

CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY

CC BOUND TO THE PROTEIN.

CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,

CC CYTOCHROME C1 AND THE RIESKE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

DR EMBL; X00017; CAA24915.1; -  
 DR InterPro; IPR000179; Cyt\_b\_b6.  
 DR Pfam; PF00032; cytochrome\_b\_c; 1.  
 DR Pfam; PF00033; cytochrome\_b\_n; 1.  
 DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
 DR PROSITE; PS00193; CYTOCHROME\_B\_QO; 1.  
 KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;  
 KW Heme; Kinetoplast.  
 FT METAL 74 IRON 1 (HEME B562 AXIAL LIGAND).  
 FT METAL 88 IRON 2 (HEME B566 AXIAL LIGAND).  
 FT METAL 175 IRON 2 (HEME B562 AXIAL LIGAND).  
 FT METAL 189 IRON 1 (HEME B566 AXIAL LIGAND).  
 SQ SEQUENCE 363 AA; 43455 MW; 2A37CD827E55923B CRC64;

Alignment Scores:  
 Pred. No.: 1.24 Length: 363  
 Score: 103.00 Matches: 45  
 Percent Similarity: 35.08% Conservativity: 22  
 Best Local Similarity: 23.58% Mismatches: 54  
 Query Match: 7.22% Indels: 70  
 DB: Gaps: 10

US-09-658-824-808 (1-781) x CYB\_TRYBB (1-363)

QY 705 GTGTTTACGCTGTCTTCAATTAACCTTGCTGCTCTTCACTGCTTC----- 655  
 Db 37 IleCysGlyValCysLeu-----AlaTrpLeuPhePheSerCysPheIleCysSer 53  
 QY 654 ---TGCATTTCACAGTGTCTCTTTAGTATTATCTTCACTTGACGCCGCCAACCCAA 598  
 Db 54 AsnTrpTyPheValLeuPheLeuTrpAspPhe-----AspLeuGlyPheVal 69  
 QY 597 ATCCAGATTATCCCGGCTGTGTGACTGATGAGCTTGTGAGATCACTTC----- 546  
 Db 70 IleArgSerValHisIleCysPheThrSerLeu-LeuTyTrpLeuLeuLeuHisIle 89  
 QY 545 -----CATGTCGGCAGCTGGATCTCAG 523  
 Db 89 ePheLysSerIleThrLeuIleLeuPheAspThrHisIleLeuValTrpPheIleG1 109  
 QY 522 CTGATATCCG---ATCTTCTTCTGTCTGCTGCCAGGTAGGATCCGACTTTTCAGCTG 466  
 Db 109 yPheIleLeuPheValPheIleIleIleAlaPheIleGlyTyValLeuProCyst 129  
 QY 465 CTGGTCTCTTTTGGGCG-----TCTCCATTAC 436  
 Db 129 hrMetMetSerTyTrpGlyLeuThrValPheSerAsnIleAlaThrValProIleL 149  
 QY 435 TGGGCTCTCG-----GACTGGCTGTGTGTCTGTCTGTGTGTG-----T 397  
 Db 149 euGlyIleTrpLeuCysTyTrpIleTrpGlySerGluPheIleAspPheThrLeuL 169  
 QY 396 GTGTCGCTCAGCCCTCCTCTCTGTCGCG----- 365  
 Db 169 euLysLeuHisValLeuHisValLeuLeuProPheIleLeuLeuIleLeuLeuLeu 189  
 QY 364 -----TCTGGGGATCCAGAACTCTCTGT----- 341  
 Db 189 isLeuPheCysLeuHisTyTrpMetSerSerAspAlaPheCysAspArgPheAlaPheT 209  
 QY 340 -----CGGGGTCCAGGTCTCTG 320  
 Db 209 yrCysGluArgLeuSerPheCysMetTrp 218  
 RESULT 10  
 DMP1\_MOUSE  
 ID DMP1\_MOUSE  
 AC O55188; STANDARD; PRT; 503 AA.

DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix  
 DE protein-1) (DMP-1) (AG1).  
 GN DMP1 OR DMP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID-10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS WEBSTER; TISSUE=Molar;  
 RX MEDLINE=96184255; PubMed=9525343;  
 RA McDougall M., Gu T.T., Luan X., Simmons D., Chen J.;  
 RT "Identification of a novel isoform of mouse dentin matrix protein 1:  
 RT spatial expression in mineralized tissues.";  
 RL J. Bone Miner. Res. 13:422-431(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=L29/SVJ; TISSUE=Spleen;  
 RA Feng J.Q., Traianedes K., Luan X., McDougall M.;  
 RT "Study of murine Dmp-1 gene function and regulation.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: COULD BE INVOLVED IN THE MINERALIZATION OF EXTRACELLULAR  
 CC MATRIX.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN  
 CC ONTOBLAST, AMEIOBLAST AND CEMENTOBLAST. ALSO EXPRESSED IN BONE  
 CC PANTICULARLY IN OSTEOBLAST.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U65020; AAB93764.1; -  
 CC EMBL; AJ242625; CAB59629.1; -  
 CC MGD; MGI:94910; Dmp1.  
 KW Extracellular matrix; Signal.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT CHAIN 17 503 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.  
 FT DOMAIN 41 44 POLY-PRO.  
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 350 352 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 35 35 D -> G (IN REF. 2).  
 FT CONFLICT 67 67 H -> D (IN REF. 2).  
 FT CONFLICT 99 99 D -> F (IN REF. 2).  
 FT CONFLICT 116 116 G -> Y (IN REF. 2).  
 FT CONFLICT 137 137 A -> T (IN REF. 2).  
 SQ SEQUENCE 503 AA; 54000 MW; 9E8AF9F2729F113A CRC64;  
 Alignment Scores:  
 Pred. No.: 2.37 Length: 503  
 Score: 99.50 Matches: 65  
 Percent Similarity: 32.03% Conservativity: 25  
 Best Local Similarity: 23.13% Mismatches: 98  
 Query Match: 7.21% Indels: 93  
 DB: Gaps: 10  
 US-09-658-824-808 (1-781) x DMP1\_MOUSE (1-503)  
 QY 22 GACTCGGGTCCCTGAGCTGGATCTTCTCCCTACTGAGACACGCCGGTAGGTCCA 81  
 Db 104 AspSerGly-----AspAspThrPheGlyAspGluAspAsnGlyLeuGlyPro 119  
 QY 82 CAGSCAGATCCAACTGGGAGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 141  
 Db 120 GluGluGlyGlnTrpGlyProSerLysLeuAspSerAspGluAspSerAla----- 137

```

QY 142 CGAGGGTTGTGTGCTGACTGACTGAGTGGAGAGCCCTCGAAGTCGTGCTGCTCCCTCTC 201
Db 138 -----AspThrGlnSerSerGlu-AspSerThrSerGlnGlnAsnSe 152
QY 202 ATCGCGTGCAGCGCCATGAGCTCTT-----GTCCTGCTCAGC 240
Db 152 rAlaGlnAspThrProSerAspSerLysAspGlnAspSerGluAspAspAlaHisSerAr 172
QY 241 GCATACCTAGG-----AGGAGGAGGCGCGAGG 270
Db 172 gProAspAlaGlyAspSerAlaGlnHisSerGluSerGluGlnArgValGlyGyl 192
QY 271 AGTGGAGGGCTCAGCGAAGCTGGGT----- 298
Db 192 ySerGluGly---GlnSerSerHisGlyAspGlySerGluPheAspAspGluGlyMetGl 211
QY 299 -----CCTGTTGGGGG 309
Db 211 nSerAspAspProGluSerThrArgSerAspArgGlyHisAlaArgMetSerSerAlaGl 231
QY 310 TATCCGAGTCCCAAGACAGCTGGA-----ACCCGACAGAGAATTCGGACT 357
Db 231 ylleArgSerGluSerLysGlyAspArgGluProThrSerThrGlnAspSerAspAs 251
QY 358 CCCAG----- 364
Db 251 pSerGlnSerValGluPheSerSerArgLysSerPheArgSerArgValSerGluGl 271
QY 365 -----CGGACAGGAGGAGCGGCGATGAG 390
Db 271 uAspTyrArgGlyGluLeuThrAspSerAsnSerArgGluThrGlnSerAspSerThrGl 291
QY 391 CGACACA---CACAAACACAGACACAGCCAGTCCAGGAGCCAGCCAAATGGAGAGC 447
Db 291 uAspThrAlaSerLysGlu-GluSerArgSerGluSerGlnGlu---AspThrAlaGlu 310
QY 448 CCATAAAGAAAGAACCCAGCAGCTGAAGTCGGATCTACACCTGGCGGACGACAGAG 507
Db 310 erGlnSerGlnGluAspSerProGluGlyGlnAspProSerSerGluSerSerGluGlu 330
QY 508 AAGATCAGGATACAGCTGAGATCCAGTCCGCGACATGGAAGTGTATCTGCAAGAGCTGC 567
Db 330 laGlyGluProSerGlnGluSerSerSerGluSerGlnGluGlyValThrSerGluSerA 350
QY 568 ATCAGTCAACACCGGGGATAATCTGGATTGGTTCGGCGCTCAAGGTGAAGATAAT 626
Db 350 rgGlyAspAsnProAspAsnThrSerGlnThrGlyAspGlnGluAspSerGluSerSer 369

RESULT 11
NR41_HUMAN STANDARD; PRT; 598 AA.
ID P22736;
AC 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Orphan nuclear receptor HMR (Early response protein NAK1) (TR3 orphan
DE receptor).
DE NR41 OR HMR OR NAK1 OR GFRP1.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=fetal skeletal muscle;
RX MEDLINE-91133413; PubMed-2283997;
RA Nakai A., Kartha S., Sakurai A., Toback F.G., Degroot L.J.;
RT "A human early response gene homologous to murine nur77 and rat
RL NGFI-B, and related to the nuclear receptor superfamily.";
RN Mol. Endocrinol. 4:1438-1443(1990).
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE-90173205; PubMed-2626032;
RA Chang C., Kokontis J., Liao S., Chang Y.;
RT "Isolation and characterization of human TR3 receptor: a member of
RT steroid receptor superfamily.";
RL J. Steroid Biochem. 34:391-395(1989).
CC -1- FUNCTION: ORPHAN NUCLEAR RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FETAL MUSCLE AND ADULT LIVER, BRAIN, AND
CC THYROID.
CC -1- INDUCTION: BY GROWTH-STIMULATING AGENTS.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR4 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D49728; BAA08565.1; -.
DR EMBL; L13740; AAA36763.1; -.
DR PIR; A37251; A37251.
DR HSSP; P19793; 2NULL.
DR TRANSFAC; T02767; -.
DR MIM; 139139; -.
DR InterPro; IPR000536; Hormone_rec_llg.
DR InterPro; IPR001723; Strdhormone_receptor.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRODHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; ZnF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Phosphorylation.
FT DNA_BIND 267 332 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 267 287 C4-TYPE.
FT ZN_FING 303 327 C4-TYPE.
FT DOMAIN 409 459 LIGAND-BINDING (POTENTIAL).
FT DOMAIN 82 92 POLY-SER.
FT DOMAIN 583 586 POLY-PRO.
FT MOD_RES 341 341 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT MOD_RES 351 351 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT CONFLICT 262 262 G -> P (IN REF. 2).
FT CONFLICT 370 370 R -> L (IN REF. 2).
SQ SEQUENCE 598 AA; 64463 MW; 41DAEA7C25FDA22 CRC64;

Alignment Scores:
Pred. No.: 2.41 Length: 598
Score: 99.50 Matches: 16
Percent Similarity: 37.11% Conservative: 56
Best Local Similarity: 28.87% Mismatches: 72
Query Match: 6.98% Indels: 50
DB: 1 Gaps: 7

US-09-658-824-808 (1-781) x NR41_HUMAN (1-598)
QY 568 TGCAGCTCTTGCAGATCACCTTCAGTTCGCGACTGGGATCTCAGCTATCCTGATCT 509
Db 78 CysSerSerAlaSerSerAlaSerSerThrSerSerSerAlaThrProGlyAla 97
QY 508 TCTTCTGTC-----TGCTGCCAGCT----- 488
Db 98 SerAlaSerPheLysPheGluAspPheGlnValTyrGlyCysTyrProGlyProLeuSer 117
QY 487 -----GTAGGATCCCGACTTTCAGCTGCTGGTCTTCTTTTGGGGCTCTCCATTA--- 437
Db 118 GlyProValAspGluAlaLeuSerSerGlySerAspTyrTyrGlySerProCysSer 137

```







\*Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons.\*;  
Cell 53:401-411(1988).  
[2]  
SEQUENCE FROM N.A.  
MEDLINE=85297791; PubMed=2994021;  
Simpson A.G., Neckelmann N., la Cruz V.F., Muhlich M.L., Simpson L.;  
"Mapping and 5' end determination of kinetoplast maxicircle gene transcripts from *Leishmania tarentolae*."; Nucleic Acids Res. 13:5977-5993(1985).  
[3]  
SEQUENCE OF 1-48 FROM N.A.  
MEDLINE=88124876; PubMed=2448777;  
Feagin J.E., Shaw J.M., Simpson L., Stuart K.;  
"Creation of AUG initiation codons by addition of uridines within cytochrome b transcripts of kinetoplastids."; Proc. Natl. Acad. Sci. U.S.A. 85:539-543(1988).  
[4]  
SEQUENCE OF 21-371 FROM N.A.  
MEDLINE=85079995; PubMed=6096360;  
de la Cruz V.F., Neckelmann N., Simpson L.;  
"Sequences of six genes and several open reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*."; J. Biol. Chem. 259:15136-15147(1984).  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL COUPLED TO ATP SYNTHESIS.  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN.  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B, CYTOCHROME C1 AND THE RIESKE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
CC -1- CAUTION: THE GENOMIC DERIVED SEQUENCE DIFFERS FROM THAT SHOWN AS IT IS MODIFIED BY EXTENSIVE RNA EDITING.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; L07542; AAA31879.1; ALT\_SEQ.  
DR EMBL; M10126; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; M19065; AAA31878.1; -;  
DR PIR; H22848; H22848.  
DR PIR; A28118; A28118.  
DR InterPro; IPR000179; Cyt\_b\_b6.  
DR Pfam; PF00032; cytochrome\_b\_c1.  
DR Pfam; PF00033; cytochrome\_b\_n1.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
DR PROSITE; PS00193; CYTOCHROME\_B\_OO; 1.  
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;  
KW Heme.  
KW METAL 82 82 IRON 1 (HEME B562 AXIAL LIGAND).  
FT METAL 96 96 IRON 2 (HEME B566 AXIAL LIGAND).  
FT METAL 183 183 IRON 2 (HEME B562 AXIAL LIGAND).  
FT METAL 197 197 IRON 1 (HEME B566 AXIAL LIGAND).  
SO SEQUENCE 371 AA; 44555 MW; 0003BD11538EA75B CRC64;  
  
Alignment Scores:  
Pred. No.: 3.6 Length: 371  
Score: 97.00 Matches: 36  
Percent Similarity: 39.47% Conservative: 24  
Best Local Similarity: 23.68% Mismatches: 46  
Query Match: 6.80% Indels: 46  
DB: 1 Gaps: 8  
  
US-09-658-824-808 (1-781) x CYB-LEITA (1-371)



CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M29297; AA088561.1; -  
DR PIR; JQ0431; JQ0431.  
KW Hypothetical protein; Transposable element.  
SQ SEQUENCE 348 AA; 35520 MW; 3BDF5D83ABBB92A CRC64;

Alignment Scores:  
Pred. No.: 4.28 Length: 348  
Score: 96.00 Matches: 40  
Percent Similarity: 34.51% Conservative: 9  
Best Local Similarity: 28.17% Mismatches: 52  
Query Match: 6.73% Indels: 41  
DB: Gaps: 5

US-09-658-824-808 (1-781) x YT35\_STRFR (1-348)

QY 348 TCTTCTGTCGGGTTCAGAGTCTTCTGGGACTCGATACCC----- 307  
Db 220 SerArgAlaGlyCysProThrAlaAlaGlySerLeuProAlaProArgProAla 239  
QY 306 -----CCACAGCAGCCACGCTT----- 289  
Db 240 SerSerAlaSerSerProGlnAlaAlaProAlaAlaProSerAlaThrArgLeuPro 259  
QY 288 ---CGCTCGAGCCCTCCACTCTCGGCCCTCTCTCTCCCTAGT----- 247  
Db 260 ArgArgThrThrProSerAlaProArgProSerSerArgProAlaArgProProlePro 279  
QY 246 TATGCCGTGACGACAGAAAGTCCATGGCGGTGGACCCGATGAGAGGACGACGAC 187  
Db 280 AlaAlaArgProProArgArgThrProGlyThrPro-Arg-----ProAlaAl 296  
QY 186 TTCGAGGCGCTTCTCACTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 127  
Db 296 aAlaArgAla-----ArgAlaProAlaGlyC 305  
QY 126 CTCTTCCTCTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 67  
Db 305 sSerProAlaArgArgThrProSerAlaProThrAspArgCysArgAlaAlaArgArg 325  
QY 66 TGTCTCAGTACGAGGAGAAATCCAGACCTCAGGACCCGAGTCCCGGCTCACAGCTC 7  
Db 325 gGlySerProArgProProAlaAlaArgProProGlyArgGlnGlyThrArgArgAspSe 345  
QY 6 CGCC 3  
Db 345 rAla 346

RESULT 16  
GGE4\_HUMAN  
ID GGE4\_HUMAN STANDARD; PRT; 117 AA.  
AC Q13068;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE GAGE-4 protein (G antigen 4).  
GN GAGE4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Melanoma;  
RX MEDLINE=95378788; PubMed=7544395;  
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,  
RA Boon T.;  
RT \*A new family of genes coding for an antigen recognized by autologous

DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS0071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
FT DOMAIN 1 140 PRO-RICH.  
FT DNA\_BIND 144 203 HOMEBOX.  
SQ SEQUENCE 277 AA; 30213 MW; BE744C143FB9F9FC CRC64;

Alignment Scores:  
Pred. No.: 3.83 Length: 277  
Score: 96.50 Matches: 39  
Percent Similarity: 39.60% Conservative: 20  
Best Local Similarity: 26.17% Mismatches: 47  
Query Match: 6.99% Indels: 43  
DB: Gaps: 8

US-09-658-824-808 (1-781) x HMPH\_CHICK (1-277)

QY 210 CCAGCCCATGGACC---TTCTTCTCTGTCAGGCCATTAATAGGAGGAGGAGGCC 266  
Db 130 ProLeuLeuTrpSerProPheIleGlnArgProLeuHisLys-----ArgLysGly 146  
QY 267 GAGGAGTGGAGCGCTCAGCGAGCTGGGTGCTGTGGGGTATCC---GAGTCCAG 323  
Db 147 GlyGlnValArgPheSerAsnGluGlnThrIleGluGluGlyLysPheGluThrGln 166  
QY 324 AAGCACCTGGAACCCGACAG---AAGATTCTG----- 353  
Db 167 LysTrpLeuSerProProGluArgLysArgLeuAlaLysLeuLeuGlnLeuSerGluArg 186  
QY 354 -----GACTCC 359  
Db 187 GlnValLysThrTrpPheGlnAsnArgArgAlaLysTrpArgArgLeuLysGlnGluAsn 206  
QY 360 CCAGACGGGACGAGGAGCGGATGAGCGACACACACACACACACACACACACAGC 419  
Db 207 ProGlnAlaThrLysLysGluGluAlaGluGlyThrGlyAspHisGlyAspProArgSer 226  
QY 420 CAG---TCCAGGAGCCAGTATGGA-----GAGCCCAAAAGAAAGAACGACCA 467  
Db 227 GluGlySerProSerProAlaGlyGlyGlyGluAlaProGlnAspSerProSerAla 246  
QY 468 GCTGAAGTGGGATCTACACCTGGGCGAGCAGACAGAAAGATCAGGATACAGCTGAG 527  
Db 247 AlaSerGlnGluAspProGluSerAspValSerAspSerAspGlnGluVal----- 264  
QY 528 ATCCAGTGGCGGACATGGAGGTGAT 554  
Db 265 -----AspIleGluGlyAsp 269

RESULT 15  
YT35\_STRFR  
ID YT35\_STRFR STANDARD; PRT; 348 AA.  
AC P20186;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 35.5 kDa protein in transposon TN4556.  
OS Streptomyces fradiae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1906;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON-Tn4556;  
RX MEDLINE=90185236; PubMed=2155856;  
RA Siemieniak D.R., Slightom J.L., Chung S.T.;  
RT \*Nucleotide sequence of Streptomyces fradiae transposable element  
TN4556: a class-II transposon related to Tn3.\*;  
RL Gene 86:1-9(1990).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```
RT cytolitic T lymphocytes on a human melanoma.*;
CC J. Exp. Med. 182:689-698(1995).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19145; AAA82747.1; .
DR MIM; 604246; .
KW Multigene family.
SQ SEQUENCE 117 AA; 12956 MW; 365ED71B2F9DC7AD CRC64;

Alignment Scores:
Pred. No.: 4.24 Length: 117
Score: 95.50 Matches: 30
Percent Similarity: 45.56% Conservative: 11
Best Local Similarity: 33.33% Mismatches: 47
Query Match: 6.92% Indels: 2
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGE4_HUMAN (1-117)
QY 411 CCACACGCGAGTCCCGAGGAGCCAGTAAATGAGAGCCCAAAAGAACACGACGCT 470
Db 29 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGluProAlaThr 48
QY 471 GAAAGTCGGGATCTTACACCTGGCGAGCAGACAGAGATCAGGATACAGCTGAGATC 530
Db 49 GlnArgGlnAspProAlaAlaAlaGlnGluGlyGluAspGluGlyAlaSerAlaGlyGln 68
QY 531 CCAGTCGCGGACATGGAAGGTGATCGAAGAGCTGCATCAGTCAACACCGGGGATAAA 590
Db 69 GlyProLysPro--GluAlaAspSerGlnGlnGlnGlyHisProGlnThrGlyCysGlu 87
QY 591 TCTGGATTGGTTCGCGGTCAAGGTGAAGATATACCTAAAGAGGAACACTGTAAAT 650
Db 88 CysGluAspGlyProAspGlyGlnGluMetAspPro-ProAsnProGluGluValLysTh 107
QY 651 GCACAGACGAGTGAAGAGCAACACAA 678
Db 107 rProGluGluGlyGluLysGlnSerGln 116

RESULT 17
GGE5_HUMAN
ID GGE5_HUMAN STANDARD; PRT; 117 AA.
AC Q13069;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-5 protein (G antigen 5).
GN GAGE5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolitic T lymphocytes on a human melanoma.*;
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19145; AAA82747.1; .
DR MIM; 604246; .
KW Multigene family.
SQ SEQUENCE 117 AA; 12956 MW; 365ED71B2F9DC7AD CRC64;

Alignment Scores:
Pred. No.: 4.24 Length: 117
Score: 95.50 Matches: 30
Percent Similarity: 45.56% Conservative: 11
Best Local Similarity: 33.33% Mismatches: 47
Query Match: 6.92% Indels: 2
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGE4_HUMAN (1-117)
QY 411 CCACACGCGAGTCCCGAGGAGCCAGTAAATGAGAGCCCAAAAGAACACGACGCT 470
Db 29 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGluProAlaThr 48
QY 471 GAAAGTCGGGATCTTACACCTGGCGAGCAGACAGAGATCAGGATACAGCTGAGATC 530
Db 49 GlnArgGlnAspProAlaAlaAlaGlnGluGlyGluAspGluGlyAlaSerAlaGlyGln 68
QY 531 CCAGTCGCGGACATGGAAGGTGATCGAAGAGCTGCATCAGTCAACACCGGGGATAAA 590
Db 69 GlyProLysPro--GluAlaAspSerGlnGlnGlnGlyHisProGlnThrGlyCysGlu 87
QY 591 TCTGGATTGGTTCGCGGTCAAGGTGAAGATATACCTAAAGAGGAACACTGTAAAT 650
Db 88 CysGluAspGlyProAspGlyGlnGluMetAspPro-ProAsnProGluGluValLysTh 107
QY 651 GCACAGACGAGTGAAGAGCAACACAA 678
Db 107 rProGluGluGlyGluLysGlnSerGln 116

RESULT 17
GGE5_HUMAN
ID GGE5_HUMAN STANDARD; PRT; 117 AA.
AC Q13069;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-5 protein (G antigen 5).
GN GAGE5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolitic T lymphocytes on a human melanoma.*;
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19146; AAA82748.1; .
DR MIM; 604247; .
KW Multigene family.
SQ SEQUENCE 117 AA; 12924 MW; 234A865E2ECDD0D6 CRC64;

Alignment Scores:
Pred. No.: 4.24 Length: 117
Score: 95.50 Matches: 30
Percent Similarity: 45.56% Conservative: 11
Best Local Similarity: 33.33% Mismatches: 47
Query Match: 6.92% Indels: 2
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGE5_HUMAN (1-117)
QY 411 CCACACGCGAGTCCCGAGGAGCCAGTAAATGAGAGCCCAAAAGAACACGACGCT 470
Db 29 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGluProAlaThr 48
QY 471 GAAAGTCGGGATCTTACACCTGGCGAGCAGACAGAGATCAGGATACAGCTGAGATC 530
Db 49 GlnArgGlnAspProAlaAlaAlaGlnGluGlyGluAspGluGlyAlaSerAlaGlyGln 68
QY 531 CCAGTCGCGGACATGGAAGGTGATCGAAGAGCTGCATCAGTCAACACCGGGGATAAA 590
Db 69 GlyProLysPro--GluAlaAspSerGlnGlnGlnGlyHisProGlnThrGlyCysGlu 87
QY 591 TCTGGATTGGTTCGCGGTCAAGGTGAAGATATACCTAAAGAGGAACACTGTAAAT 650
Db 88 CysGluAspGlyProAspGlyGlnGluMetAspPro-ProAsnProGluGluValLysTh 107
QY 651 GCACAGACGAGTGAAGAGCAACACAA 678
Db 107 rProGluGluGlyGluLysGlnSerGln 116

RESULT 18
GGE6_HUMAN
ID GGE6_HUMAN STANDARD; PRT; 117 AA.
AC Q13070;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-6 protein (G antigen 6).
GN GAGE6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolitic T lymphocytes on a human melanoma.*;
RL J. Exp. Med. 182:689-698(1995).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U19147; AA82749.1; -  
DR MIM; 604248; -  
KW Multigene family.  
SQ SEQUENCE 117 AA; 12892 MW; 234A865E3FCCCD06 CRC64;

Alignment Scores:  
Pred. No.: 4.24 Length: 117  
Score: 95.50 Matches: 30  
Percent Similarity: 45.56% Conservative: 11  
Best Local Similarity: 33.33% Mismatches: 47  
Query Match: 6.92% Indels: 2  
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGE6\_HUMAN (1-117)

QY 411 CCACACAGCCAGTCCAGGAGCCCAAGTAAATGGAGAGCCCAAAAGAAAGAACCCAGCAGCT 470  
DB 29 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGluProAlaThr 48  
QY 471 GAAAGTCGGATCCTACACCTGGCGACGACAGACAGAACATCAGGATACGCTGAGATC 530  
DB 49 GlnArgGlnAspProAlaAlaAlaGlnGluGlyGluAspGluGlyAlaSerAlaGlyGln 68  
QY 531 CCAGTCGGCGACATGGAAGGTGATCTCGAAGAGCTGCATCAACACCCGGGATATAA 590  
DB 69 GlyProLysPro---GluAlaAspSerGlnGlnGlnGlnHisProGlnThrGlyCysGlu 87  
QY 591 TCTGGATTTGGTTCGGCGCTCAAGGTGAAGATATACCTTAAGAGGAAACACTGTAAAT 650  
DB 88 CysGluAspGlyProAspGlyGlnGluValAspPro-ProAsnProGluGluValLysTh 107  
QY 651 GCCAAGACGAGTGAAGACCAACCAAA 678  
DB 107 rProGluGluGlyLysGlnSerGln 116

RESULT 19  
GGE3\_HUMAN STANDARD; PRT; 118 AA.  
AC Q13067;  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE GAGE-3 protein (G antigen 3).  
GN GAGE3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Melanoma;  
RX MEDLINE=95378788; PubMed=7544395;  
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,  
Boon T.;  
RT "A new family of genes coding for an antigen recognized by autologous  
cytolytic T lymphocytes on a human melanoma.";  
J. Exp. Med. 182:689-698(1995).  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT  
NOT IN NORMAL TISSUES, EXCEPT TESTIS.  
CC -1- SIMILARITY: BELONGS TO THE GAGE FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U19144; AA82746.1; -  
DR MIM; 604245; -  
KW Multigene family.  
SQ SEQUENCE 118 AA; 12937 MW; D97EEB19E735103 CRC64;

Alignment Scores:  
Pred. No.: 4.25 Length: 118  
Score: 95.50 Matches: 30  
Percent Similarity: 45.56% Conservative: 11  
Best Local Similarity: 33.33% Mismatches: 47  
Query Match: 6.92% Indels: 2  
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGE3\_HUMAN (1-118)

QY 411 CCACACAGCCAGTCCAGGAGCCCAAGTAAATGGAGAGCCCAAAAGAAAGAACCCAGCAGCT 470  
DB 30 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGluProAlaThr 49  
QY 471 GAAAGTCGGATCCTACACCTGGCGACGACAGACAGAACATCAGGATACGCTGAGATC 530  
DB 50 GlnArgGlnAspProAlaAlaAlaGlnGluGlyGluAspGluGlyAlaSerAlaGlyGln 69  
QY 531 CCAGTCGGCGACATGGAAGGTGATCTCGAAGAGCTGCATCAACACCCGGGATATAA 590  
DB 70 GlyProLysPro---GluAlaAspSerGlnGlnGlnGlnHisProGlnThrGlyCysGlu 88  
QY 591 TCTGGATTTGGTTCGGCGCTCAAGGTGAAGATATACCTTAAGAGGAAACACTGTAAAT 650  
DB 89 CysGluAspGlyProAspGlyGlnGluMetAspPro-ProAsnProGluGluValLysTh 108  
QY 651 GCCAAGACGAGTGAAGACCAACCAAA 678  
DB 108 rProGluGluGlyLysGlnSerGln 117

RESULT 20  
MY16\_MOUSE STANDARD; PRT; 657 AA.  
ID MY16\_MOUSE  
AC P17564;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Myeloid differentiation primary response protein MyD116.  
GN MYD116.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=SL;  
RC MEDLINE=90251472; PubMed=2339071;  
RX Lord K.A., Hoffman-Liebermann B., Liebermann D.A.;  
RT "Sequence of MyD116 cDNA: a novel myeloid differentiation primary  
RL Nucleic Acids Res. 18:2823-2823(1990).  
CC -1- INDUCTION: BY INTERLEUKIN-6.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; X51829; CAA36128.1; -  
DR PIR; S10001; S10001.  
DR MGD; MGI:1100516; Myd116.  
KW Differentiation; Antigen; Repeat.  
FT DOMAIN 283 451 4.5 X TANDEM REPEATS.

```
FT REPEAT 283 322 1.
FT REPEAT 323 360 2.
FT REPEAT 361 398 3.
FT REPEAT 399 436 4.
FT REPEAT 437 451 5 (INCOMPLETE).
SQ SEQUENCE 657 AA; 71840 MW; 9B217001019C38A7 CRC64;

Alignment Scores:
Pred. No.: 5,4 Length: 657
Score: 95.00 Matches: 61
Percent Similarity: 35.77% Conservative: 27
Best Local Similarity: 24.80% Mismatches: 93
Query Match: 6.88% Indels: 65
DB: 1 Gaps: 12

US-09-658-824-808 (1-781) x MY16_MOUSE (1-657)
QY 55 GCTACTGACACAGCGGGGTAGTCCACAGGAGAGTCCAACTGGAGTTGAAGTGTGAGT 114
Db 212 AlaThrGluGluGlyThrGluAsnLysAlaAspProSerAsnSerProSer 229
QY 115 GAGAGTGAAGAGAACACAGCAGGCTCCGGAGGGTGTGTGCTGACTCAGAGTGAG 174
Db 230 -----SerGlySerHisSerArgAlaTrpGluTyrTyrSerArgGlu 243
QY 175 AAGGCCCTCGAAGTCTGCTCCTCTCATGCGGTGCCAGCCCATGGACCTTCTTGTCTC 234
Db 244 LysProLysGlnGluGlyGluAlaLysValGluAlaHisArgAlaGlyGlnGlyHisPro 263
QY 235 GTCACGGCCATAACTAGGAGGAGGAGGAGGCGGAGG-----AGTGGAGGG 279
Db 264 CysArgAsnAlaGluAlaGluGlyGlyProGluThrPheValCysThrGlyAsn 283
QY 280 GCTCAGGCGAAGTGGGTGCTGTGGGGGTATCCGAGTCCAGAGCAGCCTGGAACCCC 339
Db 284 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 296
QY 340 GACAGAAATCTGAGTCC-----CCAGAGCGGACACAGGAGGAGGAGGAGGAGGAG 393
Db 296 pThrGluGluGlu-AspAsnSerAspSerAspSerAlaGluGluAspThrAla 313
QY 394 CACACACAAACAGAACACACAGCCAGTCC-----ACACTGGGCGAGC 498
Db 314 --GlnThrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 333
QY 426 -----CAGAGCCCACTAATGGAGAGCCCAAAAGAGAACACAGCAGCTG 471
Db 333 LysGluAspThrGluGluGluAspSerAspSerAlaGluGluAspThrAlaGlnT 353
QY 472 AAAGTCGGGATCCCT-----ACACTGGGCGAGC 498
Db 353 hrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyGlu 373
QY 499 AGACAGAGAGATCAGATACACTGATGCCAGTCCGCGGACATGGAAGGTGATCTGC 558
Db 373 spThrGluGluGluAsnSer-----AspLeuAspSerAlaGluG 386
QY 559 AAGAGCTGCATCAGTCAAC-----ACCGGGGATAAATCTGGATT-----GGGTTC 606
Db 386 LuAspThrAlaGlnThrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValT 406
QY 607 GCGCTCAAGTGAAGATTAACCTAAAGAGAACACTGTAAATGCCAGAGCAGGTGAA 666
Db 406 YrArgProGlyGluAsp-----Thr-GluGluGluAsnSerAspLeuAspSerAlaGluGlu 424
QY 667 GAGCAACCAACAA 678
Db 425 AspThrAlaGln 428

RESULT 21
ID YFHA_BORPE STANDARD; PRT; 320 AA.
AC P33445;
```

```
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 33.8 kDa protein in PHAC 3' region (ORFA).
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RA Willems R.J.L.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X64876; CAA46093.1; -
CC Hypothetical protein.
KW SEQUENCE 320 AA; 33781 MW; 1B5C54B11DBF2D73 CRC64;

Alignment Scores:
Pred. No.: 7,91 Length: 320
Score: 92.50 Matches: 45
Percent Similarity: 37.79% Conservative: 20
Best Local Similarity: 26.16% Mismatches: 65
Query Match: 6.70% Indels: 42
DB: 1 Gaps: 8

US-09-658-824-808 (1-781) x YFHA_BORPE (1-320)
QY 122 AAGAGAACACAGCAGCTCCGAGGTTGTGTGTCAG-----TGACTC 166
Db 142 ArgArgAsnGlnGlnAlaGlnGlyThrGlyAlaGlyGlnArgAlaGluAspLeuValLeu 161
QY 167 AGACTGAGAGGCCCTCGAAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 226
Db 162 Argile-AlaAlaProArgGlnPheGlyAspGlyHis----- 173
QY 227 CTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 286
Db 174 -LeuAlaHisArgGlyAlaGly-----GlyGlyArgArgAlaGlyAsn---ArgGlyGluAs 191
QY 287 CGAAGCTGGGGTCTGTGGGGGTATCCGAG-----TCCAGAAAGCACCTGGAACCCGACA 343
Db 191 pGlyAlaAlaAspValGlyValGlnGlnProAlaGlyGlnProValGlnProArgAr 211
QY 344 GAAGATTCTGGACTCCCGAGCGGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 403
Db 211 gGluAlaLeuGluGlnValLeuGlyGlnAlaGlyAlaGlyGlnAspLeuAlaHisProGl 231
QY 404 CACAGAACACACAGCAGCTCCAGGAGCCAGTAATGGAGAGCCCAAAAGAA-----458
Db 231 nGluGlnArgGlnGlyGlnArgProAlaArgGlyAlaProAspArgAspGlyH 251
QY 459 -----GAACGACGCTGAAAGTCCGGATCTCTACACCTGGGCGAG-----497
Db 251 sGlyValAlaGlyArgProAlaGlyGluAspPheHisAlaAspProGlyArgAlaGlyGl 271
QY 498 -----CAGACAGAGAGATCAGATACAGCTGAGAT 529
Db 271 nGlyGluAlaAsnProAspSerAlaThrGlnGlnArgGluAspArgAsp-----287
QY 530 CCCAGTCCGCGACATCGAAGGTGATCTGCAAGAG 563
Db 288 -----AspGlnGlnGlyAspAspGlnAsp 295

RESULT 22
ID YFHA_BORPE STANDARD; PRT; 320 AA.
AC P33445;
```

```
ID IB63_HSV2H STANDARD; PRT; 512 AA.
AC P28276;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcriptional regulator IE63 (VMM63) (ICP27).
GN UL54.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10315;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92113549; PubMed=1662697;
RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT *Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2.*
RT J. Gen. Virol. 72:3057-3075(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE DOWN-REGULATION OF VIRAL IMMEDIATE-EARLY
CC GENES AND ACTS IN COMBINATION WITH ICP0 AND ICP4 AS AN ACTIVATOR
CC OF LATE GENES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,
CC HSV-2 UL54, EBV-1 5, VZV 4, EBV BMLF1, HCMV UL69, AND HVS-1 57.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D10471; BAA01269.1; -.
DR EMBL: Z85099; CAB06702.1; -.
DR PIR: JQ1498; WBEXA.
KW Early protein; Transcription regulation; DNA-binding.
SQ SEQUENCE 512 AA; 54958 MW; 459651470A503BA7 CRC64;

Alignment Scores:
Pred. No.: 8.24 Length: 512
Score: 92.50 Matches: 35
Percent Similarity: 42.50% Conservative: 16
Best Local Similarity: 29.17% Mismatches: 40
Query Match: 6.70% Indels: 29
DB: 1 Gaps: 6

US-09-658-824-808 (1-781) x IE63_HSV2H (1-512)
QY 177 GGCCCTCGAAGTCGTCTG-----CCCTCTCATCGCGTG 209
Db 140 GlyArgArgGlyArgArgArgGlyArgGlyArgTyrGlyProGlyGlyAlaAspSerThr 159
QY 210 CCACGCCCATGGACCTTCTGCTCTGTCAGCGGCATCACTAGGAGGAGGAGGCGCGAG 269
Db 160 ProlysProArgArgArgValSerArgAsnAlaHisGlnGlyGlyArgHisProAla 179
QY 270 GAGTGGAGGCGCTCAGCGGAAGCTGGGGTGTCTGTGGGGGTATCCGAGTCCCAAGACAC 329
Db 180 SerAlaArgThrAspGlyProGlyAlaThr-----His 190
QY 330 CTGGAACCCGACAG-----AAGATTCTGGACTCCCGACAGCGGACACAGG 380
Db 191 GlyGluAlaArgArgGlyGlyGluGlnLeuAspValSerGlyGlyProArgProArgGly 210
QY 381 ACGCATGAGCG-----ACACACAAACACACAGAACACACAGCAGCTCC 425
Db 211 ThrArgGlnAlaProProProLeuMetAlaLeuSerLeuThrProHisAlaAspGly 230
```

```
QY 426 CAGGAGCCCGCAGTATGGAGAGCGCCCAAAAGAAAGAACACAGCAGCTCAAACTCGGATCCT 485
Db 231 ArgAlaPro-----ValProGluArgLysAlaProSerAlaAspThrIleAspPro 247
RESULT 23
GGE2_HUMAN STANDARD; PRT; 116 AA.
ID GGE2_HUMAN
AC Q13066;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-2 protein (G antigen 2).
GN GAGE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Melanoma;
RX van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RX Boon T.;
RT *A new family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.*;
RT J. Exp. Med. 182:689-698(1995).
CC -1- FUNCTION: ANTIGEN, RECOGNIZED ON MELANOMA BY AUTOLOGOUS CYTOLYTIC
CC T LYMPHOCYTES.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -1- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U19143; AAA82745.1; -.
DR EMBL: 604244; -.
DR MIM: 604244; -.
KW Multigene family; Antigen.
SQ SEQUENCE 116 AA; 12786 MW; DD305D5CA29AF19A CRC64;

Alignment Scores:
Pred. No.: 7.9 Length: 116
Score: 92.00 Matches: 37
Percent Similarity: 38.97% Conservative: 16
Best Local Similarity: 27.21% Mismatches: 60
Query Match: 6.67% Indels: 23
DB: 1 Gaps: 4

US-09-658-824-808 (1-781) x GGE2_HUMAN (1-116)
QY 273 TGAAGGGGCTCAGCGGAAGCTGGGGTGTCTTGGGGGTATCCGAGTCCCAAGACACCTG 332
Db 3 TrpArgGlyArgSerThrTyrArg-----ProArgProArgArgTyrVal 17
QY 333 GAACCCCGACAGAAGATTCTGGACTCCCGACAGCGGACAGGAGGAGCGCATGAGCG 392
Db 18 GluProGluMetIle-----GlyProMetArg----- 27
QY 393 ACACACAAACACACAGAACCCACACAGCAGCTCCCGAGGCCATATGAGAGCCCCAA 452
Db 28 -----ProGluGlnPheSerAspGluValGluProAlaThrProGlu 41
QY 453 AAGAGAAACACACAGCTGAAAGTGGGATCTACACCTGGGCGAGCAGCAGAGAGAT 512
Db 42 GluGlyGluProAlaThrGlnArgGlnAspProAlaAlaGlnGlyGluAspGlu 61
QY 513 CAGGATACACCTGAGATCCCGAGTCCGCGACATGGAAGGTGATCTGCAAGAGCTGCATCAG 572
```

Db 62 GlyAlaSerAlaGlyGlnGlyProLysPro---GluAlaHisSerGlnGluGlnGlyHis 80  
 QY 573 TCAACACCGGGGATAAATCGGATTTCGGGTCCGGCTCAAGGTGAAGATAAATACCTAA 632  
 Db 81 ProGlnThrGlyCysGluCysGluAspGlyProAspGlyGlnGluMetAspPro-ProAs 100  
 QY 633 AGAGGAACACTGTAAATGCCAGAGAGGTGAAGAGCAACCAACAA 678  
 Db 100 nProGluGluValLysThrProGluGluGlyGlnLysGlnSerGln 115

RESULT 24  
 TAU\_PAPHA  
 ID TAU\_PAPHA STANDARD; PRT; 382 AA.  
 AC Q9MYX8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Microtubule-associated protein tau (Neurofibrillary tangle protein)  
 DE (Paired helical filament-tau) (PHF-tau).  
 GN MAPT OR TAU.  
 OS Papio hamadryas (Hamadryas baboon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopitheciinae; Papio.  
 OX NCBI\_TaxID=9557;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Frontal cortex;  
 RA Wang X.L., Wang J., Schultz C., Hubbard G.B.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: PROMOTES MICROTUBULE ASSEMBLY AND STABILITY, AND MIGHT  
 CC BE INVOLVED IN THE ESTABLISHMENT AND MAINTENANCE OF NEURONAL  
 CC POLARITY. THE C-TERMINUS BINDS AXONAL MICROTUBULES WHILE THE N-  
 CC TERMINUS BINDS NEURAL PLASMA MEMBRANE COMPONENTS, SUGGESTING THAT  
 CC TAU FUNCTIONS AS A LINKER PROTEIN BETWEEN BOTH. AXONAL POLARITY IS  
 CC DETERMINED BY TAU LOCALIZATION (IN THE NEURONAL CELL). IN THE  
 CC DOMAIN OF THE CELL BODY DEFINED BY THE CENTROSOME (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: MOSTLY FOUND IN THE AXONS OF NEURONS, IN THE  
 CC CYTOSOL AND IN ASSOCIATION WITH PLASMA MEMBRANE COMPONENTS (BY  
 CC SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.  
 CC -!- DOMAIN: THE TAU/MAP REPEAT BINDS TO TUBULIN.  
 CC -!- PTM: PHOSPHORYLATION AT VARIOUS SERINE AND THREONINE RESIDUES IN  
 CC S-P OR T-P MOTIFS BY PROLINE-DIRECTED PROTEIN KINASES (PDPK: CDC2,  
 CC CDK5, GSK3, MAPK) (A FEW SITES PER PROTEIN IN INTERPHASE, MORE IN  
 CC MITOSIS), AND AT SERINE RESIDUES IN K-X-G-S MOTIFS BY  
 CC MAP/MICROTUBULE AFFINITY-REGULATING KINASE (MARK) (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 4 TAU/MAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF281310; AAF97596.1;  
 DR InterPro; IPR002955; Tau\_protein.  
 DR InterPro; IPR001084; Tubulin-bind.  
 DR Pfam; PF00418; tubulin-binding; 4.  
 DR PRINTS; PR01261; TAUPROTEIN.  
 DR PROSITE; PS00229; TAU\_MAP; 4.  
 KW Microtubules; Cytoskeleton; Repeat; Acetylation; Phosphorylation.  
 FT INIT\_MET 0  
 FT REPEAT 185 215 TAU/MAP MOTIF 1.  
 FT REPEAT 216 246 TAU/MAP MOTIF 2.  
 FT REPEAT 247 277 TAU/MAP MOTIF 3.  
 FT REPEAT 278 309 TAU/MAP MOTIF 4.  
 FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 FT DISULFID 232 263 BY SIMILARITY.  
 SQ SEQUENCE 382 AA; 39879 MW; D2D15A53AA00E87 CRC64;

Alignment Scores:  
 Pred. No.: 8.78 Length: 382  
 Score: 92.00 Matches: 59  
 Percent Similarity: 40.38% Conservative: 25  
 Best Local Similarity: 28.37% Mismatches: 79  
 Query Match: 6.67% Indels: 45  
 DB: 1 Caps: 11

US-09-658-824-808 (1-781) x TAU\_PAPHA (1-382)

QY 76 GGTCCACAGGCAGATCAACTGGGAGTTGAAGTGTGAGTGAAGAGGAACCAACGCA 135  
 Db 41 GlyLeuLysAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAlaAla 60  
 QY 136 GGTTCCTCGGAGGGTGTGTGCTGACTCAGTCAAGTGAAGAGCCCTCGAAGTGTGCTG 195  
 Db 61 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 74  
 QY 196 CCTCTCATGCGTCCACGCCCTGACCTTCTGTCTGTCACGGCCATACTAGGAG 255  
 Db 74 p-----GlyThrGlySerAsp-----As 80  
 QY 256 GAAGAGGGGCGGAGGAGTGGGGCTCAGGCGA----- 289  
 Db 80 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 100  
 QY 290 -----AGTGGGTGCTGTGGGGGTATCCGAGTC-----CCAGAAGCAAC 330  
 Db 100 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProAlaPr 120  
 QY 331 TGGAAACCCACACAGAAATCTGGACTCCACAGAGCGGACAGAGAGGAGCGCATGAG 390  
 Db 120 oLysThrPro---ProSerGlyGluProLysSerGlyAspArgSerGlyTyrSe 139  
 QY 391 CGACACACACAAACACA---GAACACACACAGCCAGTCCAGAGCCAGTAATGGAGAGC 447  
 Db 139 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 159  
 QY 448 CCAAAAAGAAGAACCA-----GCACCTGAAGTCGGGATCCTACACCTGGCGCAGAGA 501  
 Db 159 roProAlaArgGluProLysLysValAlaValArgThrProProLysSerProSerS 179  
 QY 502 CAGAAGAAGATCAG--GATACAGCTCAGATCCAGTCCGCGACATGGAAGGTGATCTGC 558  
 Db 179 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 197  
 QY 559 AAGAGCTGCATCAGTCAACACCGGGGATAAATCTGATTTGGTTCGCGCGTCAAGGTG 618  
 Db 198 LysSerLysIleGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly---LysVal 216  
 QY 619 AAGATAATACCTAAAGAG 636  
 Db 217 GlnIlelleAsnLysLys 222

RESULT 25  
 TISD\_HUMAN  
 ID TISD\_HUMAN STANDARD; PRT; 482 AA.  
 AC P47974;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Butyrate response factor 2 (TIS1LD protein) (EGF-response factor 2)  
 DE (ERF-2).  
 GN BRF2 OR TIS1LD OR ERF2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96132724; PubMed=8545129;  
 RA Ino T., Iisui H., Hirano M., Kurosawa Y.  
 RT \*Identification of a member of the Tis1l early response gene family

at the insertion point of a DNA fragment containing a gene for the T-cell receptor beta chain in an acute T-cell leukemia.";  
 Oncogene 11:2705-2710(1995).

[2]  
 SEQUENCE FROM N.A.  
 MEDLINE=95137407; PubMed=7835719;  
 Nie X.F., Maclean K.N., Kumar V., McKay I.A., Bustin S.A.;  
 "ERF-2, the human homologue of the murine Tisild early response gene";  
 Gene 152:285-286(1995).  
 CC -1- FUNCTION: PROBABLE REGULATORY PROTEIN INVOLVED IN REGULATING THE  
 CC RESPONSE TO GROWTH FACTORS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: CONTAINS 2 C3H1-TYPE ZINC FINGERS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

EMBL; U07802; AAA91778.1; -;  
 DR EMBL; X78992; CAA55592.1; -;  
 DR InterPro; IPR000571; zf-CCCH.  
 DR Pfam; PF00642; zf-CCCH; 2.  
 DR SMART; SM00356; Znf-C3H1; 2.  
 DR Nuclear protein; Repeat; Metal-binding; Zinc-finger; DNA-binding.  
 KW ZN\_FING 159 178 C3H1-TYPE 1.  
 FT ZN\_FING 197 216 C3H1-TYPE 2.  
 FT DOMAIN 106 109 POLY-GLY.  
 FT DOMAIN 138 141 POLY-GLN.  
 FT DOMAIN 143 146 POLY-GLY.  
 FT DOMAIN 288 291 POLY-PRO.  
 FT DOMAIN 323 330 POLY-ALA.  
 FT DOMAIN 382 388 POLY-ALA.  
 FT DOMAIN 393 399 POLY-GLN.  
 FT CONFLICT 96 97 DL -> TS (IN REF. 2).  
 FT CONFLICT 318 318 A -> T (IN REF. 2).  
 FT CONFLICT 329 330 AA -> LR (IN REF. 2).  
 FT CONFLICT 450 450 S -> SLSDRDSYLSG (IN REF. 2).  
 SQ SEQUENCE 482 AA; 49866 MW; 72335B1C92FD706E CRC64;

## Alignment Scores:

Pred. No.: 8 96 Length: 482  
 Score: 92.00 Matches: 60  
 Percent Similarity: 36.44% Conservative: 22  
 Best Local Similarity: 26.67% Mismatches: 83  
 Query Match: 6.45% Indels: 60  
 DB: 1 Gaps: 11

US-09-658-824-808 (1-781) x TISD\_HUMAN (1-482)

QY 637 CCTCTTAGTATATCTTCACCTTCAGCGCGGAACCAATCCAGATTATCCCGGTG 578  
 ||||| :||| ||||| ||| ||| ||| |||  
 Db 277 ProLeu---LeuLeuAspSerProThrSerArgThrProPro-----ProSer 292  
 QY 577 TTTGACTGATGACGCTTCGACATCACCTTCATCGCGACATGGGATCTACGTGTA 518  
 ||||| ||||| ||| :||| ||| :|||  
 Db 293 CysSerSerAlaSerCysSerSerAlaSerSerCysSerSerAlaSerAlaAla 312  
 QY 517 TCTGATCTCTTCT-----GTCCTGCTGCCAGGTGTAGGATCCCGACTTTCAGCTGCT 464  
 ||| ||| ||||| ||| ||| ||| |||  
 Db 313 SerThrProSerGlyAlaProThrCysCys-----AlaSerAlaAlaAlaAla 329  
 QY 463 GGTCTCTCTTTT-----GGGCTCTCCATTACTGGCTCCGCGGACGTGGCTGTGT 413  
 ||||| :||| ||||| ||||| ||||| |||||  
 Db 330 -AlaLeuLeuGlyThrGlyGlyAlaGluAspLeuLeuAlaProGlyAlaProCysAl 349  
 QY 412 GGTCTGTGTGTGTGCTCATCGCGGCCCTCTCCTGGTCCGCTGGGAGTCC 353  
 ||| ||| ||| :||| :||| |||  
 Db 349 aAla-----CysSerSerAlaSerCysAlaAsnAsnAlaPheAlaPheGlyProGl 366

QY 352 AGAATCTTCTCGGGGTTCCAGGTGCTTCTGGGACTCGGATACCCCAACAGCACCCCA 293  
 ||||| :||| ||||| ||| ||| ||| |||  
 Db 366 uLeuSerSerLeu-----1leThrProLeuAla1leGl 377  
 QY 292 GCTTCGCTGAGCCCTCCACTCTCGGCCCTCTCTTCTCCCTAGTATGCGCGTGACGA 233  
 | :||| :||| :||| :||| :||| :||| :|||  
 Db 377 nThrHisAsnPheAlaAlaValaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 394  
 QY 232 GACAAGAAGTCCATGGCGGTGGCAGCCATGAGAGGACGACGACTTCGAGGCGCTTCT 173  
 : :||| :||| :||| :||| :||| :||| :|||  
 Db 394 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 403  
 QY 172 CACTCTGAGTCACTGACACACACACACACACCTCCGGAAGCCTGCTGGTCTCTTCTCACTCTCAC 113  
 ||| ||| ||| ||| ||| ||| |||  
 Db 404 -----ProAlaGlnProPro-AlaProProSerAlaThr----- 414  
 QY 112 TCACACTTCACTCCAGTTGGATCTGCCTGTGGACCTACC----- 72  
 ||||| ||||| ||||| ||||| ||||| |||||  
 Db 415 -----LeuProAlaGlyAlaAlaAlaProProSerProPheSerPheGlnL 431  
 QY 71 -----CCCGCTGTCTCAGTAGCGGAGAAAGAAATCCAGACCTCAGGACCCGAGTCGCGG 17  
 ||||| :||| :||| ||| ||| ||| |||  
 Db 431 euProArgArgLeuSerAspSerProValPheAspAlaProProSerProProAspSerS 451  
 QY 16 CTCACAGCTCC 6  
 : :|||||  
 Db 451 erLeuSerSer 454  
 RESULT 26  
 NR41\_CANFA  
 ID NR41\_CANFA STANDARD; PRT; 598 AA.  
 AC P51666;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Orphan nuclear receptor HMR (Orphan nuclear receptor NGFI-B).  
 GN NR4A1 OR HMR.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97050607; PubMed=8895335;  
 RA Pichon B., Jimenez-Cervantes C., Pirson I., Maenhaut C.,  
 RA Christophe D.;  
 RT "Induction of nerve growth factor-induced gene-B (NGFI-B) as an ea-ly  
 RT event in the cyclic adenosine monophosphate response of dog  
 RT thymocytes in primary culture.";  
 RL Endocrinology 137:4691-4698(1996).  
 CC -1- FUNCTION: ORPHAN NUCLEAR RECEPTOR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR4 SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC EMBL; X97226; CAA65863.1; -;  
 DR HSP; P19793; 2NLL.  
 DR InterPro; IPR000536; Hormone\_rec\_lig.  
 DR InterPro; IPR001723; Sterdormone\_receptor.  
 DR InterPro; IPR001628; zf-C4.  
 DR Pfam; PF00104; hormone\_rec; 1.  
 DR Pfam; PF00105; zf-C4; 1.  
 DR PRINTS; PR00398; STRDHORMONER.  
 DR PRINTS; PR00047; STROIDFINGER.



DR	SMART; SM00430; HOL1; 1.	
DR	SMART; SM00399; ZNF_C4; 1.	
DR	PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.	
DR	Receptor; Transcription regulation; DNA-binding; Nuclear protein;	
KW	zinc-finger; Phosphorylation.	
FT	DNA_BIND	267 332
FT		NUCLEAR RECEPTOR-TYPE.
FT	2N_FING	267 287
FT		C4-TYPE.
FT	2N_FING	303 327
FT		C4-TYPE.
FT	DOMAIN	409 459
FT		LIGAND-BINDING (POTENTIAL).
FT	DOMAIN	82 92
FT		POLY-SER.
FT	DOMAIN	583 586
FT		POLY-PRO.
FT	MOD_RES	341 341
FT		PHOSPHORYLATION (BY PKA) (BY SIMILARITY)
FT	MOD_RES	351 351
FT		PHOSPHORYLATION (BY PKA) (BY SIMILARITY)
SO	SEQUENCE	598 AA; 64432 MW; 773563A8EC48F906 CRC64;

Alignment Scores:					
Pred. No.:	9.14	Length:	598		
Score:	92.00	Matches:	52		
Percent Similarity:	32.6%	Conservative:	15		
Best Local Similarity:	25.3%	Mismatches:	64		
Query Match:	6.45%	Indels:	74		
DB:	1	Gaps:	11		

US-09-658-824-808 (1-781) x NR41\_CANFA (1-598)

Qy	568	TGCAGCTCTTGCACATCACCTTCCTCCATGTCGGCACCACGTCTCAGCTATCATCTGATCT	509
Db	78	CysSerSerAlaSerSerSerAlaSerSerThrSerSerSerSerAlaThrSerProAla	97

QY 508 TC<sup>1</sup>CTGTC-----TGCTGCCAGGT-----488  
 |||:::  
 |||:::

Db 98 SerAlaSerPheLysPheGluAspPheGlnValItyrGlycystyrProGlyProLeuSer II

Db 118 GlyProLeuAspGluThrLeuSerSerGlySerAspTyrTyrGlySerProCysSer 137

Qy 436 -----CTGGGCTCCTGGGACTGGCTG 416  
||| ||| ||| ||| |||

[illegible]

Db 158 PheGly-----  
          |||  
          ProPheSerPro 163

QY 355 TCCAGAATCTTCTGTGCGGGTTCCAGGTGCTTCTGGGACTCGGATACCCCCCAAC----- 302

Db	164	SerGlnThrTyr---GlucGlyLeuArgAla---	TrpIleHrgLucInLeuProLysAlaSer	187
Qa:	201	-----	NCCNCCCGCAGCCmGCCCGCCmC	275

Db 182 GlyHisProGlnProAlaPheSerProThrGlyProSerPro--- 200

Qy 274 CACTCCTCGGCCTCCTTCTCTCCCTAGTTATGGCCGTACGAGACAGAAGGTCATGGG 215

DB ---SerLeuAlaGlnSerProLeuLysLeuPnePro-----Ser 211  
201  
214 CCTCCACACCCACACACACACACACACACACACACACACACAC 155  
Q41 CCTCCACACCCACACACACACACACACACACACACACACACAC 155

Db 213 GlnAlaThrCysGlnLeuGlyGluArg-----GluSerTyrSerIleSer 227

QY 154 ACACACCCCTCGGAAGCCTGCTGGTTCTCTTCACTCTCACTTCACACTTCAACTCCCAG 95

04	nncccccccccccc	80
226	tnrtatpneprogly-----	LeuAIApIotInrSeIFtHISLeuAspGlyProGlyMet
244		

Db	246	LeuAspAlaProVal	250

RESULT 27  
SM6C\_RAT

ID	SMOC_RAT	STANDARD;	PKI;	500 AA.
AC	Q9WTL3; Q9WTM6;			

```

30-MAY-2000 (Rel. 39, Created)
30-MAY-2000 (Rel. 39, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Semaphorin 6C precursor (Semaphorin Y) (Sema Y).
GN Sema6C OR SEMA Y.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Muscle;
RA MEDLINE=99160821; PubMed=10049528;
RX Kikuchi K., Chedotal A., Hanafusa H., Ujilima Y., de Castro F.,
RA Goodman C.S., Kimura T.;
RA "Cloning and characterization of a novel class VI semaphorin,
ET semaphorin Y."
RL MOL. Cell. Neurosci. 13:9-23(1999).
CC -1- FUNCTION: SHOWS GROWTH CONE COLLAPSING ACTIVITY ON DORSAL ROOT
CC GANGLION (DRG) NEURONS IN VITRO. MAY BE A STOP SIGNAL FOR THE DRG
CC NEURONS IN THEIR TARGET AREAS, AND POSSIBLY ALSO FOR OTHER
CC NEURONS. MAY ALSO BE INVOLVED IN THE MAINTENANCE AND REMODELING OF
CC NEURONAL CONNECTIONS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SEMA Y-L (SHOWN HERE) AND SEMA
CC Y-S; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY REGIONS OF THE DEVELOPING
CC NERVOUS SYSTEM, PROBABLY IN NEURONS AND THEIR PRECURSORS, BUT ALSO
CC IN NONNEURAL TISSUE SUCH AS IMMATURE MUSCLE AND DERMIS. IN ADULT,
CC STRONG EXPRESSION IN THE SKELETAL MUSCLE AND MODERATE EXPRESSION
CC IN THE BRAIN, WHERE CEREBELLUM SHOWS THE HIGHEST EXPRESSION. ALSO
CC EXPRESSED IN ALMOST ALL AREAS OF THE CNS.
CC -1- DEVELOPMENTAL STAGE: DETECTED AT E12 AND FOUND AT MARKEDLY
CC INCREASED LEVEL AT E15 AND E18 IN BOTH THE HEAD AND THE BODY. AT
CC BIRTH THE LEVEL DECREASES SIGNIFICANTLY.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000817; BAA76293.2; -
DR EMBL; AB014074; BAA76295.1; -
DR InterPro; IPR001627; Sema.
DR Pfam; PF01403; Sema; 1.
KW Signal; Transmembrane; Multigene family; Neurogenesis; Glycoprotein;
KW Developmental protein; Alternative splicing.
FT SIGNAL 1 23
FT CHAIN 24 960 SEMAPHORIN 6C.
FT DOMAIN 24 635 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 636 656 POTENTIAL.
FT DOMAIN 657 960 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 234 541 SEMA.
FT DOMAIN 693 699 POLY-PRO.
FT DOMAIN 783 786 POLY-PRO.
FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 586 617 MISSING (IN ISOFORM SEMA Y-S).
SQ SEQUENCE 960 AA; 102610 MW; C88293C5607E6086 CRC64;

Alignment Scores:
Pred. No.: 9.53 Length: 960
Score: 92.00 Matches: 61
Percent Similarity: 32.39% Conservative: 19
Best Local Similarity: 24.70% Mismatches: 72
Query Match: 6.45% Indels: 96
DB: 1 Gaps: 13

```





Db	649	LysGlyGlyPhe - PheSerSerPheMetLysLysArgAsnAlaPro	-----ThrProProLys 667
QY	96	TGCGAGTTGAAGTGTGAGTGAAGAGGAACCA	-----GC 134
Db	667	sArgSerSerPheArgGluMetGluAsnGlnProHisLysLysTyrcIuLeuThrG1	-----
QY	135	AGGCTTCGGAGGTTGTG	-----TGCTCAGTCACTCAGAGTGA 173
Db	687	yAsnPheSerValAlaSerLeuGlnHisAlaAspLysPheSerPheThrProAlaG1	707
QY	174	GAAGGCCCTCGAAGTCGTCCTCTCATCGGTGCCCGCCATCGGACCTTCCTGTCT	233
Db	707	ngInGluAlaasnLeuValPro	-----LysCys-----TyrglycylSerPheAl 723
QY	234	CTCTACGCCCACTAGGGAGGAAGGAGGCCGAGGAGTGGAGGGCTCAGGCGAAGCT	293
Db	723	agInArgAsnLeucyCysAsnAspGlyGly	-----GlyglyglycylSerGlyThrAl 741
QY	294	GGSGTCTGTGG	-----GGGTAT----- 312
Db	741	adglyglytyrpsrGlyIlethrGlyPhePheThrProArgLeuIleLysLysThrIe	761
QY	313	-----	-----CCGAGTCCCAAG 326
Db	761	uGlyLeuArgAlaGlyLysProThrAlaSerAspAspThrSerLysPheProPheArgSe	781
QY	327	CACCTGGAAACCCGACAGACATCTCGACTCCCGACAGCGGACAC	----- 373
Db	781	rasn - SerThrSerSerMetSerGlyLeuProGluGlnAspArgMetAlaMetThrL	801
QY	374	-----	-----GAGAGGACGCCATGAGCGCACAC 398
Db	801	euProArgAsnCysGlnArgSerLysLeuGlnLeuGluArg - Thr	-----ValSer 817
QY	399	ACAAACACAGAACACACACGCCAGTCCCGAGGCCAGTAATGGAGACCCCAAAA	----- 455
Db	818	ThrSerSerGlnProGluAsnValAspArgAlaAsnAspMetLeuProLysLysSer	837
QY	456	GAAGAACACAGACGTGAAAGTCGGGATCTTACACCTGGCGCAGACAGAGAAGACATCAG	515
Db	838	GluGluSerAlaAlaProSerArgGluArgProLysAlaLysLeuLeu	-----ProArg 855
QY	516	GATACAGCTGATCCAGTCGCGGACATGGAAAGTGATCTGCAAGAGTGTCATCAGTCA	575
Db	856	GlyAlaThrAlaLeuProLeuArgThrProSerGlyAspLeuAlaIleThrGluLys - As	875
QY	576	AACACCGGGATAAATCTGCATTTGGTTCCGGCGTCAAGGTGAAGATAAATACCTAAAGA	635
Db	875	pProGlyValcylVal	-----AlaGlyValAlaAlaProLysGlyLysG1 892
QY	636	GAACAC	-----TGTAATAATGCCAAGCAGGTGAAGAG 669
Db	892	uLysAsnGlyGlyAlaArgLeuGlyMetAlaGlyValProGluAspGlyGluGln	910
RESULT 29			
OTX1_HUMAN			
ID	OTX1_HUMAN	STANDARD;	PRT; 354 AA.
AC	P32242;		
DT	01-OCT-1993 (Rel. 27, Created)		
DT	01-OCT-1993 (Rel. 27, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	Homeobox protein OTX1.		
GN	OTX1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Carnathini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93327763; PubMed=8101484;		
RA	Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,		
RA	Nigro V., Boncinelli E.;		
RT	*A vertebrate gene related to orthodenticle contains a homeodomain of		

RT	the bicoid class and demarcates anterior neuroectoderm in the gastrulating mouse embryo.*;
RL	EMBO J. 12:2735-2747(1993).
CC	- - FUNCTION: PROBABLY PLAY A ROLE IN THE DEVELOPMENT OF THE BRAIN AND THE SENSE ORGANS. CAN BIND TO THE BCD TARGET SEQUENCE (BTS); 5'-TCTAATCCC-3'.
CC	- - SUBCELLULAR LOCATION: Nuclear (Probable).
CC	- - TISSUE SPECIFICITY: EXPRESSED IN BRAIN.
CC	- - DEVELOPMENTAL STAGE: EMBRYO.
CC	- - SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS. "BICOID" SUBFAMILY.
DR	HSSP; P22808; 1VND.
DR	TRANSFAC; T02079; -.
DR	MIM; 600036; -.
DR	InterPro; IPR001356; Homeobox.
DR	InterPro; IPR003025; Otx_TF.
DR	Pfam; PF00046; homeobox; 1.
DR	PRINTS; PR00024; HOMEBOX.
DR	PRINTS; PR01255; OTXHOMEBOX.
DR	SMART; SM00389; HOX; 1.
DR	PROSITE; PS00027; HOMEBOX_1; 1.
DR	PROSITE; PS00071; HOMEBOX_2; 1.
KW	Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT	DNA_BIND 38 97 HOMEBOX.
FT	DOMAIN 275 301 HIS-RICH.
SQ	SEQUENCE 354 AA; 37327 MW; E32C1E4746BDAA CRC64;
 Alignment Scores:	
Pred. No.:	9.53
Score:	Length: 354
Percent Similarity:	Matches: 50
Best local Similarity:	Conservative: 20
Query Match:	Mismatches: 59
DB:	Indels: 63
	Gaps: 10
 US-09-658-824-808 (1-781) x OTXL_HUMAN (1-354)	
Qy	622 TCCTCACCTTGAGCGCGGAACCAATCCAGATTATCCCCGGTGTTCAGCATGCAGC 563              : : :
Db	112 SerSerProValargGlusSerGlySerGluSerGlyGlnPhe----- 127
Qy	562 TCTTGCAATCACCTTCCATGTGCGGCATCGGATCTCAGCTGATCTGATCTCTTTCT 503              : : :
Db	128 -----ThrProAlaValSerSerAlaSerSerSerAlaSerSerSer 145
Qy	502 GTCTGCTGCCAGGTGTAGATCCCGACTTCACGTGCTGTTCTCTTTTGGGCTCT 443     : : :       : : :
Db	146 SerAlaasnProala-----AlaAlaAlaAlaAlaGlyLeuGlyGlyAsn 160
Qy	442 CCATTACTGGGCTCTCTGGGACTGGCTGTGTTGTGTGTCGCTCATGCC 383 
Db	161 Pro-----ValAlaAlaAla 165
Qy	382 GTCCTCTCCTGGTCCGCTCTGGGAGTCCAGAATCTTCTGTGGGGTTCAGGTCCTTC 323 
Db	166 SerSerLeuSerThrProAlaAlaSerSer-----Ile 176
Qy	322 TGGGAC-----TCGGATACCCCAACACACCACCCAGCTTCG-----CCTGAGGCC 278         : : :
Db	177 TrpSerProAlaSerIleSerProGlySerAlaProAlaSerValSerValProGluPro 196
Qy	277 CTC-----CACTCCTCGGCCCTCTCTCCTCCCTAGTTATGGCGTAGCAGAC 230     : : :       : : :
Db	197 LeuAlaAlaProSerAsnThrSerCysMet-----GlnArg 208
Qy	229 AAGAAGGTCATGGCGGTGGCAGCGCATGAGAGGGACGACGACTT-----CGAGGG 179 
Db	209 SerValAlaAlaGlyAlaAlaThrAlaAlaAlaAlaSerTyroProMetSerTyroGlyGlnGly 228
Qy	178 CCTTCTCACTGTGACTGACTGACACACACACCCCTCC----- 143     : : :
Db	229 GlySerTyroGlyGlnGlyTyroProTherProSerSerSerTyroPheGlyValAspCys 248     : : :

QY 142 GGAGCCTCGTTCCTTCTCACTCTCACTCACAC 107  
 Db 249 SerSerTyrLeuAlaProMethHisSerHisHis 260  
 RESULT 30  
 DMP1\_RAT  
 ID DMP1\_RAT STANDARD; PRT; 489 AA.  
 AC P98193;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix  
 DE protein-1) (DMP-1) (AG1).  
 GN DMP1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Tooth;  
 RX MEDLINE=93286101; PubMed=8509401;  
 RT \*Characterization of a novel dentin matrix acidic phosphoprotein.  
 RT \*Characterization of a novel dentin matrix acidic phosphoprotein.  
 RT J. Biol. Chem. 268:12624-12630(1993).  
 CC -1- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF  
 CC EXTRACELLULAR MATRIX.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN ODONTOBLAST  
 CC AND AMELOBLAST.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L11354; -; NOT\_ANNOTATED\_CDS.  
 KW Extracellular matrix; Signal.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT CHAIN 17 489 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.  
 FT DOMAIN 41 44 POLY-PRO.  
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 334 336 CELL ATTACHMENT SITE (POTENTIAL).  
 SQ SEQUENCE 489 AA; 53058 MW; 59F8381479DDA085 CRC64;  
 Alignment Scores:  
 Pred. No.: 14 Length: 489  
 Score: 89.50 Matches: 68  
 Percent Similarity: 34.25% Conservative: 32  
 Best Local Similarity: 23.29% Mismatches: 105  
 Query Match: 6.49% Indels: 87  
 DB: 1 Gaps: 12  
 US-09-658-824-808 (1-781) x DMP1\_RAT (1-489)  
 QY 22 GACTCGGTCCTCGAGGTCTGGATCTTCTCCGCTACTGAGACACGGCGGTAGTGCCA 81  
 Db 89 AspSerGly-----AspAspThrPheGlyAspGluAspAsnGlyProGlyPro 104  
 QY 82 CAGGCAGATCCAACTGGGAGTGTGAAGTGTGAGTGTGAGAGTGAAGACACGACGCTTC 141  
 Db 105 GluGluArgGlnTrpGlyGlyProSerArgLeuAspSerAspGluAspSerAla----- 122  
 QY 142 CGGAGGTTGTGTCGTGTCAGTCACTCACTGAGACAGCCCTCAAGTCGTCCTCTC 201  
 Db 123 -----AspThrThrSerSerGlu-AspSerThrSerGlnGluAsnSe 137  
 QY 202 ATGGGTCGCCACGCCCATGGACCTTCCT-----GTCTGTCGACGGCC 243

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AB013729; BAA76294.1; -  
 CC MGD: MGI:1338032; Sema6c.  
 CC InterPro: IPR001627; Sema.  
 CC Pfam: PF01403; Sema; 1.  
 CC Signal: Transmembrane; Multigene family; Neurogenesis; Glycoprotein;  
 CC Developmental protein.  
 CC SIGNAL 1 25  
 CC CHAIN 26 931  
 CC DOMAIN 26 605  
 CC TRANSMEM 606 626  
 CC POTENTIAL.  
 CC POTENTIAL.  
 CC CYTOPLASMIC (POTENTIAL).  
 CC SEMA.  
 CC POLY-PRO.  
 CC POLY-PRO.  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 71 71  
 CC CARBOHYD 287 287  
 CC CARBOHYD 438 438  
 CC SEQUENCE 931 AA; 99537 MW; B0D99D594209F125 CRC64;

## Alignment Scores:

Pred. No.: 14.8 Length: 931  
 Score: 89.50 Matches: 60  
 Percent Similarity: 32.23% Conservative: 18  
 Best Local Similarity: 24.79% Mismatches: 70  
 Query Match: 6.28% Indels: 95  
 DB: 1 Gaps: 13

US-09-658-824-808 (1-781) x SM6C\_MOUSE (1-931)

Qy 638 TCCTTTAGTATATCTTACCTTACCGCGGACCCAAATCCAGATTATCCCGGT 579  
 Db 595 SerAlaTyRGLyValArgAspLeuSerProAlaSerAlaSerArgSerIlePro--- 603  
 Qy 578 GTTTCAGTATGACGCTCTTGACAGATCCATCTCCATGTCGCGCACTGGGATCTCAGCTGT 519  
 Db 604 ---1leProLeuLeu--- 610  
 Qy 518 ATCTGATCTCTCTCTGCTGCCAGGTAGGATCCCGACTTTCAGCTCTGGTTC 459  
 Db 611 ValAlaAlaAlaPheAla--- 616  
 Qy 458 TTTCTTTGGGGCTCTCCATTACTGGGCTCTGGGACTGCTGTGTTCTGTTGT 399  
 Db 617 ---LeuGlyAlaSerValSerGlyLeuLeu---ValSerCys 628  
 Qy 398 GTGTGCTGCTCAGC---CCT 381  
 Db 629 AlaCysArgAlaAlaAsnArgArgSerLysAspIleGluThrProGlyLeuProArg 648  
 Qy 380 CCCTC-----TCCTGGTCCCGCTG-----GGAGTCCAGAA--- 349  
 Db 649 ProLeuSerLeuArgSerLeuAlaArgLeuHisGlyGlyProGluProProProPro 668  
 Qy 348 -----TCTTCTCTCGGGGTCCAGGTGCTTCTGGGACTCGGATACCCCA--- 304  
 Db 669 ProLysAspGlyAspAlaAlaGlnThrProGlnLeuTyThrThrPheLeuProProPro 688  
 Qy 303 -----ACAGACCCCGAGCTCGCTGACGCCCTCCAGCTCTCGGGCCCTCTCTCTC 253  
 Db 689 AspGlyGlySerProProGluLeuAlaCysLeuProThrProGluThrThrProGluLeu 708

Qy 252 CCTAGTTATGGCGGTGACGAGACAAGAGGTCCATGGCGTGGCCGATGAGAGGAC 193  
 Db 709 Pro-VallYsHisLeuArgAlaSerGlyGlyProTrpGluTrpAsnGln---AsnGlyAs 727  
 Qy 192 GACGACTTCGAGGGCTCTTCTACTCTGAGTCTACCAACCCCTCCGGAACCTGTC 133  
 Db 727 nAsnAlaSerGluGlyProGly--- 738  
 Qy 132 TGGTTCCTCTTCTACTCTCTACTCTACACACTTCAACTCCAGTTCGCTGGACCTAC 73  
 Db 738 gGlyCysSer-----GlyAlaGlyGlyProAl 747  
 Qy 72 CGCGGTGCTCTAGTCGCGGAGAAAGATCCAGACTTCAGGACCGAGTCCCGGTCTCA 13  
 Db 747 aproArgValLeuVal-----ArgProProProGlyCysProGlyG 762  
 Qy 12 CAGC 9  
 Db 762 nAla 763

## RESULT 32

MAN1\_MOUSE STANDARD; PRT; 331 AA.  
 ID MAN1\_MOUSE Q9WU40;  
 AC Q9WU40;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Inner nuclear membrane protein Man1 (Fragment).  
 GN MAN1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20138223; PubMed=10671519;  
 RA Lin F., Blake D.L., Callebaut I., Skerjanc I.S., Holmer L.,  
 RA McBurney M.W., Paulin-Levasseur M., Worman H.J.;  
 RT "MAN1, an inner nuclear membrane protein that shares the LEM domain  
 RL J. Biol. Chem. 275:4840-4847(2000).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER NUCLEAR  
 CC MEMBRANE.  
 CC -1- SIMILARITY: CONTAINS 1 LEM DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF112300; AAD31594.1; -  
 CC InterPro: IPR003887; LEM.  
 CC Pfam: PF03020; LEM; 1.  
 CC SMART: SM00540; LEM; 1.  
 CC Nuclear Protein; Transmembrane.  
 CC DOMAIN 8 51  
 CC LEM.  
 CC DOMAIN 52 62  
 CC POLY-GLN.  
 CC DOMAIN 74 79  
 CC POLY-ASN.  
 CC DOMAIN 151 157  
 CC POLY-GLY.  
 CC DOMAIN 264 272  
 CC POLY-GLU.  
 CC NON\_TER 331 331  
 CC SEQUENCE 331 AA; 34594 MW; BB7AD1BD85FCF54D CRC64;

## Alignment Scores:

Pred. No.: 14.8 Length: 331  
 Score: 89.00 Matches: 51  
 Percent Similarity: 39.23% Conservative: 20  
 Best Local Similarity: 28.18% Mismatches: 62  
 Query Match: 6.45% Indels: 48  
 DB: 1 Gaps: 9









Best Local Similarity: 18.75% Mismatches: 62  
Query Match: 6.38% Indels: 72  
DB: 1 Gaps: 6

US-09-658-824-808 (1-781) x NEUM\_SERCA (1-241)

QY 223 CCTTCTGCTCGTCACGCCCACTAGGAGGAGGAGGCGGAGGAGTGGAGGGCT 282  
DB 54 ProAlaSerGluThrAspAlaAlaAspLysGluGluGlyProAlaGlyGlyAlaAla 73  
QY 283 CAGCGGAGAGCTGGGTGCTGTGGGGTATCCGAGTCCCAAGAACCTGGAACCCG-- 340  
DB 74 GluAsnLys-----Glu-SerGluAlaProAlaAlaThrG1 85  
QY 341 -----ACAGAATTCTGGACCTCCCGAG 363  
DB 85 uAlaAlaAlaAlaAspSerAlaGlnGluGluGlySerLysAspSerAlaProAl 105  
QY 364 ACGGACAGGAGGAGGAGGATGATGAGGACACACACAAACACAGACCCAGT 423  
DB 105 aGluGluLysLysGlyAspGlyAlaAlaAspThr-GlySerGluGlnProAlaProGlnA 125  
QY 424 CCAGGAGCCAGTAAATGAGAG-----CCCCAAAAG 456  
DB 125 IaAlaThrProAlaAlaSerSerGluGluLysThrAlaAlaAlaAlaProGluArgG 145  
QY 457 AAGAACAGCAGCTGAAGTCGGATCTACACCTGGCGGAGCAGACAGAAAGATCAGG 516  
DB 145 LuSerThrProLysAlaSerThrAspAsnSerProSerLysAlaAspGluAlaGlnA 165  
QY 517 AT-----ACAGCTGAGATCCAGTCGGCAG 542  
DB 165 spLysGluGluProLysGlnAlaAspValProAlaAlaAspThrThrAlaThrThrP 185  
QY 543 -----ATGGAAGGTGATCTGCAAGAGCTGCATC 570  
DB 185 roAlaAlaGluAspAlaThrAlaLysAlaThrAlaGlnProGlnMetGluThrValGlu 205  
QY 571 AGTCAACACCCGGGTAATCTGGATTGGGTTCGGCGTCAAGGTGAAGATAATACCT 630  
DB 205 erSerGlnThrGluGluLysThr----- 212  
QY 631 AAAGAGGAACACTGTAAATGCCAGAGCAGGTGAAGACCAACCAAGATTAAATGAAG 690  
DB 213 -----AspAlaValGluGluThrLysProThrGluSerAlaGlnGlnG 227  
QY 691 ACAAGCTGAACCAACCAAG 710  
DB 227 LuGlu4eLysGluGluGlu 233

RESULT 37  
TAU\_MACMU  
ID TAU\_MACMU STANDARD; PRT; 458 AA.  
AC P57786;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Microtubule-associated protein tau (Neurofibrillary tangle protein)  
DE (Paired helical filament-tau) (PHF-tau).  
GN MAPT OR TAU  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
OC Cercopitheciinae; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D).  
RC TISSUE=Brain;  
RX MEDLINE=97012131; PubMed=8858947;  
RA Nelson P.T., Stefansson K., Gulcher J., Saper C.B.;  
RT "Molecular evolution of tau protein: implications for Alzheimer's  
RT disease.";  
RL J. Neurochem. 67:1622-1632(1996).

-! FUNCTION: PROMOTES MICROTUBULE ASSEMBLY AND STABILITY. AND MIGHT BE INVOLVED IN THE ESTABLISHMENT AND MAINTENANCE OF NEURONAL POLARITY. THE C-TERMINUS BINDS AXONAL MICROTUBULES WHILE THE N-TERMINUS BINDS NEURAL PLASMA MEMBRANE COMPONENTS, SUGGESTING THAT TAU FUNCTIONS AS A LINKER PROTEIN BETWEEN BOTH. AXONAL POLARITY IS PREDETERMINED BY TAU LOCALIZATION (IN THE NEURONAL CELL) IN THE DOMAIN OF THE CELL BODY DEFINED BY THE CENTROSOME. THE SHORT ISOFORMS ALLOW PLASTICITY OF THE CYTOSKELETON WHEREAS THE LONGER ISOFORMS MAY PREFERENTIALLY PLAY A ROLE IN THE STABILIZATION. IN THE SUBCELLULAR LOCATION: MOSTLY FOUND IN THE AXONS OF NEURONS, IN THE CYTOSOL AND IN ASSOCIATION WITH PLASMA MEMBRANE COMPONENTS. -! ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; TAU A (SHOWN HERE), TAU-B, TAU-C AND TAU-D; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER FROM EACH OTHER BY THE PRESENCE OR ABSENCE OF UP TO 4 EXONS. ONE OF THESE OPTIONAL EXONS CONTAINS THE ADDITIONAL TAU/MAP REPEAT. THE SEQUENCE SHOWN HERE IS THAT OF THE COMPLETE ISOFORM NOT YET CHARACTERIZED.

-! TISSUE SPECIFICITY: EXPRESSED IN NEURONS.

-! DOMAIN: THE TAU/MAP REPEAT BINDS TO TUBULIN. TYPE I ISOFORMS CONTAIN 3 REPEATS WHILE TYPE II ISOFORMS CONTAIN 4 REPEATS.

-! PTM: PHOSPHORYLATION AT VARIOUS SERINE AND THREONINE RESIDUES IN S-P OR T-P MOTIFS BY PROLINE-DIRECTED PROTEIN KINASES (PDK: CDC2, CDK5, GSK3, MAPK) (A FEW SITES PER PROTEIN IN INTERPHASE, MORE IN MITOSIS), AND AT SERINE RESIDUES IN K-X-G-S MOTIFS BY MAP/MICROTUBULE AFFINITY-REGULATING KINASE (MARK) (BY SIMILARITY).

-! SIMILARITY: CONTAINS 4 TAU/MAP REPEATS.

InterPro: IPR002955; Tau.protein.  
InterPro: IPR001084; Tubulin-bind.  
Pfam: PF00418; tubulin-binding; 4.  
PRINTS: PR01261; TAU/PROTEIN.  
DR PROSITE; PS00229; TAU\_MAP; 4.  
DR Microtubules; Cytoskeleton; Repeat; Alternative splicing; Acetylation; Phosphorylation.  
KW INIT\_MET 0 BY SIMILARITY.  
FT REPEAT 261 291 TAU/MAP MOTIF 1.  
FT REPEAT 292 322 TAU/MAP MOTIF 2.  
FT REPEAT 323 353 TAU/MAP MOTIF 3.  
FT REPEAT 354 385 TAU/MAP MOTIF 4.  
FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
FT DISULFID 308 339 BY SIMILARITY.  
FT VARSPPLIC 44 101 MISSING (IN ISOFORM TAU-B AND ISOFORM TAU-C).  
FT VARSPPLIC 185 202 MISSING (IN ISOFORM TAU-C AND ISOFORM TAU-D).  
FT VARSPPLIC 292 322 MISSING (IN ISOFORM TAU-B).  
SQ SEQUENCE 458 AA; 47841 MW; 68BE63DBA30665A8 CRC64;

Alignment Scores:  
Pred. NO.: 18.2 Length: 458  
Score: 88.00 Matches: 57  
Percent Similarity: 36.44% Conservative: 25  
Best Local Similarity: 25.33% Mismatches: 82  
Query Match: 6.38% Indels: 61  
DB: 1 Gaps: 10

US-09-658-824-808 (1-781) x TAU\_MACMU (1-458)

QY 76 GGTCCACAGGCAGATCCCAACTGGGAGTTGAAGTGTGACTGAGAGTGAAGAGGACCAACCA 135  
DB 99 GlyThrThrAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAlaAla 118  
QY 136 GCCTTCCGAGGGTGTGTGTCAGTCACTGAGTGAAGAGCCCTCGAAGTCTGCTGTC 195  
DB 119 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 132  
QY 196 CCTCTCATCGGTGCCAGCCCATGGACCTTCTTCTCTCGTCACGCCCACTAGGAG 255  
DB 132 p-----GlyThrGlySerAsp-----As 138  
QY 256 GAAGAGGAGGCCAGAGGTGGAGGGCTCAGCGCA----- 289  
DB 138 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 158



```
QY 290 -----AGCTGGGGTCTGTTGGGGGTATCCGAGTCCCA----- 322
Db 158 oProGlycInLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProProAlaPr 178
QY 323 -----GAAGCACTGGAAACCC 339
Db 178 oLysThrProProSerAlaThrLysGlnValGlnArgLysProProProAlaGluPr 198
QY 340 GACAGACATTCGGACTCCCGACCGGACGAGGAGGAGCGGATGAGCGACACACA 399
Db 198 oThrSerGluArgGlyGluProProLysSerGlyAspArgSerGlyTrSer-SerProG 218
QY 400 CAACACAC---GAACACACAGCAGTCCCGAGGAGCGAGTAAATGGAGAGCCCAAAAG 456
Db 218 lySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrProProAla 238
QY 457 AAGAACCA-----GCAGCTGAAGTGGGATCCTACACCTGGCGACGACAGAGAAG 510
Db 238 rgGluProLysLysValAlaValArgThrProProLysSerProSerSerAlaLys 258
QY 511 ATCAG---GATACAGCTGAGATCCCGAGTGGCGACATGGAAGTGATCTGCAAGAGCTGC 567
Db 258 erArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal---LysSerLys 276
QY 568 ATCAGTCAAAACCGGGGATTAATCTGGATTGGGTTCCGGCTCAAGGTGAAGATAATA 627
Db 277 ileGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly---LysValGlnIle 295
QY 628 CCTAAAGAG 636
Db 296 AsnLysLys 298

RESULT 38
G19P_BOVIN
ID G19P_BOVIN STANDARD; PRT; 533 AA.
AC Q28034;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Protein kinase C substrate, 60.1 kDa protein, heavy chain (PKCSH)
DE (80K-H protein).
GN PKCSH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HOLSTEIN; TISSUE=Corpus luteum;
RA Brule S., Rabahi F., Beckers J., Monniaux D., Silversides D.W.,
RA Lussier J.G.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U49178; AAA92060.1; -
CC HSSP; P02593; IAK6.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR000886; ER_target.
CC InterPro: IPR002172; LDL_recept_A.
CC SMART; SM00192; LDLA; 1.
CC PROSITE; PS00014; ER_TARGET; 1.
CC PROSITE; PS00018; EF_HAND; 1.
CC Phosphorylation.
FT DOMAIN 313 334 ASP/GLU-RICH (ACIDIC)..
SQ SEQUENCE 533 AA; 60151 MW; 507858677PB81E01 CRC64;
```

```
Alignment Scores:
Pred. No.: 18.4 Length: 533
Score: 88.00 Matches: 46
Percent Similarity: 35.75% Conservative: 28
Best Local Similarity: 22.22% Mismatches: 77
Query Match: 6.38% Indels: 56
DB: 1 Gaps: 7
```

US-09-658-824-808 (1-781) x G19P\_BOVIN (1-533)

```
QY 79 CCACAGCCACATCAACTGGGAGTTGAAGTGTGAGTCAGAGTGAAGAG----- 126
Db 237 ProGluLeuAspThrAspGlyAspGlyAlaLeuSerGlyGlyGlnAlaGlnThrLeuLeu 256
QY 127 -----GAACACACAGAGGCTTCGGAGGGTGTGTGTCAGTCAGTCAGTCAGTC 168
Db 257 GlyGlyAspAlaGlnMetAspAlaAlaPhePheThrAspArgValTrpAlaAlaIleArg 276
QY 169 AGTGAGNAGCCCTCGAAGTCGTCGTCCTCTCTATGCGGTGCCACGCCCATGGACCTCT 228
Db 277 AspLysTyrrArgSerGluValLeuProThr-----GluTyrrProProSer 291
QY 229 TGTCTCGTCACGGCCATACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 288
Db 292 ProProAlaProAspVal-HetGluProLysGluGluGln----- 304
QY 289 AAGCTGGGGTGTGTTGGGGGTATCCGAGTCCCAAGAGCACCTGGAAACCCCGACAGAAGA 348
Db 305 -----ProProMetProSerProProThrGluGlu 314
QY 349 TTCTGGACTCCCGACAGCGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 408
Db 314 uGluAspGluAspGluGluAspGluGluThrGlu-----GluAspGluAspGluGlu-A 332
QY 409 AACCCACACAGCCAGTCCCGAGGAGCCAGTAATGGA-----GAGCCCAAA 453
Db 332 spGluAspSerGlnGlyGluGlnProLysAspAlaProProProAlaProAlaProGlnT 352
QY 454 AAGAAGAACCCAGCGTGAAGTCCGATCTACACCTGGCGACAGACAGACAGAA----- 506
Db 352 hrAlaSerProThrGluGluAspArgMetProProTyrrAspGluGlnThrGlnAlaPheI 372
QY 507 -----GAAGATCAGATACAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTC 543
Db 372 leAsnAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGluArgSerLeuLysAspM 392
QY 544 TGAAGGTGATCTCAAGAGAGTCATCAGTCAAAACACCGGGGATAAATCTGGATTTGGT 603
Db 392 etGluGluSerIleArgAsnLeuGluGln-----GluIleSerPheAsp 407
QY 604 TCCGGCGTCAAGGTGAA 620
Db 407 heGlyProAsnGlyGlu 412
```

RESULT 39

```
ID CIKE.DROME STANDARD; PRT; 1174 AA.
AC Q02280;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Potassium channel protein eag.
GN EAG.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91262635; PubMed-1840699;
RA Warmke J., Drysdale R., Ganetzky B.;
```



QY 339 CGACAGAAGATTCTGGACTCCCCCAG

Search completed: October 18, 2002, 09:41:08  
Job time : 42 secs

Alignment Scores:			
Pred. No.:	19.8	Length:	1208
Score:	88.00	Matches:	58
Percent Similarity:	34.67%	Conservative:	20
Best Local Similarity:	25.78%	Mismatches:	94
Query Match:	6.38%	Indels:	53
DB:	1	Gaps:	10
US-09-658-824-808 (1-781) x RCQ4_HUMAN (1-1208)			
Qy	66	ACGGCGGGTAGGTGCCAGGCAGATCCAACTGGGAGTGTGAAGTGTGACTGAGAGTGAAGA	125
Db	47	ThrLeuLysargThrThrGlyGlnAlaGlyGlyLeu-	Arg 60
		-----	
Qy	126	GGAACACGAGCGTTCGGAGGGTTGT---GTGGTCAGTGCATCAGATGAGTGAAGAGCCCT	182
		:::	:::
		:::	:::
Db	61	SerSerGluSerLeuProAlaAlaAlaGluAlaProGluProArgCysTrpGlyPro	80
		-----	
Qy	183	CGAAGTCGTCCCTCATCGGGTGCACGCCCATGGACCTTCTTGTCGTGCACGGC	242
Db	81	HisLeuAsnargAlaAlaThrLysSerProGlnProThrProGlyArgSerArgGlnGly	100
		-----	
Qy	243	CAT-----AACTAGGGAAGAGGGCCGAGGAGTGGAGGGGTTCAGGGCAAGCTGGG	296
		:::	:::
		:::	:::
Db	101	SerValProAspTyrGlyGlnArgLeuLysAlaAsnLeuLysGlyThrLeuGlnAlaGly	120
		-----	
Qy	297	GTGCTGTTGGG-----GATCCGACTCCAGAGGACCTGGNACC	338
Db	121	ProAlaLeuGlyArgArgProTrpProLeuGlyArgAlaSerSerLysAlaSerThrPro	140
		-----	
Qy	339	CGACAGAAGATTCTGGACTCCCCAGACGGGACCCAGGAGGACGGCATGAGCGACACAC	398



PF 30-JUN-2000; 2000WO-US18061.  
XX  
PR 30-JUN-1999; 99US-0346492.  
PR 15-OCT-1999; 99US-0419356.  
PR 17-DEC-1999; 99US-0466867.  
PR 30-DEC-1999; 99US-0476300.  
PR 06-MAR-2000; 2000US-0519642.  
PR 22-MAR-2000; 2000US-0533077.  
PR 10-APR-2000; 2000US-0546259.  
PR 27-APR-2000; 2000US-0560406.  
PR 05-JUN-2000; 2000US-0589184.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
PI Retter MW, Mannion J;  
XX WPI; 2001-071488/08.  
XX  
PT Lung tumor-associated proteins and the nucleic acids that encode them,  
PT useful for preventing, diagnosing and treating lung cancer -  
XX  
PS Example 1; Page 427; 436pp; English.  
XX  
CC The present invention describes immunogenic portions of lung tumour-  
CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
CC (I) have cytostatic activity and can be used in gene therapy, antisense  
CC inhibition and in vaccines. The NAS and the lung tumour-associated  
CC proteins they encode may be used in the prevention, treatment and  
CC diagnosis of diseases associated with their inappropriate expression,  
CC especially lung cancers. For example, the NAS may be administered to  
CC treat diseases by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of the protein by expressing inactive proteins  
CC or to supplement the patient's own production of (I). Additionally, the  
CC NAS may be used to produce the lung-tumour associated protein, according  
CC to standard recombinant DNA methodology. Conversely, antisense NA  
CC molecules may be administered to down regulate protein expression by  
CC binding with the cells own genes and preventing their expression. The NA  
CC and complementary sequences may also be used as DNA probes in diagnostic  
CC assays to detect and quantitate the presence of similar NA sequences in  
CC samples, and hence which patients may be in need of treatment for lung  
CC cancer. The (I) may be used as antigens in the production of antibodies  
CC and in assays to identify modulators (agonists and antagonists) of the  
CC expression and activity of the protein. AAF68083 to AAF68878 and  
CC AAB76848 to AAB76878 represent human lung tumour protein related  
CC nucleotide and protein sequences which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 160 AA:  
Alignment Scores:  
Pred. No.: 6,7e-81 Length: 160  
Score: 887.00 Matches: 160  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 64.28% Indels: 0  
DB: 22 Gaps: 0  
US-09-658-824-808 (1-781) x AAB76877 (1-160)  
QY 202 ATGCGGTGCCAGCCGACCTTCTTGTCTCGTCACGGCATAAATAGGAGGAAGGA 261  
DB 1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrArgGluGluGly 20  
QY 262 GGGCCGAGAGTGGAGGGCTCAGCGAGCTGGGGTGTGTGGGGGTATCCGAGTCCC 321  
DB 21 GlyProArgSerGlyValAlaGlnAlaLysLeuGlyCysCysTrpGlyTyrProSerPro 40  
QY 322 AGAAGCACCCTGAACCCGACAGAAGATTCTTGACTCCCGACGGGACGAGGAGGA 381  
DB 41 ArgSerThrTrpAsnProHisArgPheThrProGlnThrGlyProGlyGluGly 60  
QY 382 CGGCATGACGACACACAAACACAGAACACACAGCCAGTCCCGAGGAGCCAGTAATG 441

Db 61 ArgHisGluArgHisThrGlnThrGlnAsnHisThrAlaSerProArgSerProValMet 80  
QY 442 GAGAGCCCCAAAAGAAAGAACACAGCAGCTGAAAGTCCGGATCCCTACACCTGGGCAGCAGA 501  
Db 81 GluSerProLysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100  
QY 502 CAGAAGATCAGGATACAGCTGAGATCCAGTCCGCGACATGGAAGGTGATCTGCAAG 561  
Db 101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysValIleCysLys 120  
QY 562 AGCTGCATCAGTCAACACACGGGATAAATCTGCATTGGGTTCCGGCTCAAGGTGAAG 621  
Db 121 SerCysIleSerGlnThrProGlyIleAsnLeuAspLeuGlySerGlyValLysValLys 140  
QY 622 ATAATACCTAAAGAGGAACTGTAAATGCCAGAACGAGGTGAAGACCAACACAGATT 681  
Db 141 IleIleProLysGluGluHisCysLysMetProGluAlaGlyGluGlnProGlnVal 160  
RESULT 2  
AAB76869  
ID AAB76869 standard; Protein; 126 AA.  
XX  
AC AAB76869;  
XX  
XX 12-APR-2001 (first entry)  
XX Human lung tumour protein related protein sequence SEQ ID NO:791.  
DE  
XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;  
KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;  
KW cytostatic; antisense inhibition.  
XX  
OS Homo sapiens.  
XX  
XX WO200100828-A2.  
XX  
XX 04-JAN-2001.  
XX  
XX 30-JUN-2000; 2000WO-US18061.  
XX  
PR 30-JUN-1999; 99US-0346492.  
PR 15-OCT-1999; 99US-0419356.  
PR 17-DEC-1999; 99US-0466867.  
PR 30-DEC-1999; 99US-0476300.  
PR 06-MAR-2000; 2000US-0519642.  
PR 22-MAR-2000; 2000US-0533077.  
PR 10-APR-2000; 2000US-0546259.  
PR 27-APR-2000; 2000US-0560406.  
PR 05-JUN-2000; 2000US-0589184.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
XX Retter MW, Mannion J;  
XX WPI; 2001-071488/08.  
XX  
XX Lung tumor-associated proteins and the nucleic acids that encode them,  
XX useful for preventing, diagnosing and treating lung cancer -  
XX  
XX Example 1; Page 413; 436pp; English.  
XX  
CC The present invention describes immunogenic portions of lung tumour-  
CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
CC (I) have cytostatic activity and can be used in gene therapy, antisense  
CC inhibition and in vaccines. The NAS and the lung tumour-associated  
CC proteins they encode may be used in the prevention, treatment and  
CC diagnosis of diseases associated with their inappropriate expression,  
CC especially lung cancers. For example, the NAS may be administered to  
CC treat diseases by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of the protein by expressing inactive proteins  
CC or to supplement the patient's own production of (I). Additionally, the  
CC NAS may be used to produce the lung-tumour associated protein, according  
CC to standard recombinant DNA methodology. Conversely, antisense NA  
CC molecules may be administered to down regulate protein expression by  
CC binding with the cells own genes and preventing their expression. The NA  
CC and complementary sequences may also be used as DNA probes in diagnostic  
CC assays to detect and quantitate the presence of similar NA sequences in  
CC samples, and hence which patients may be in need of treatment for lung  
CC cancer. The (I) may be used as antigens in the production of antibodies  
CC and in assays to identify modulators (agonists and antagonists) of the  
CC expression and activity of the protein. AAF68083 to AAF68878 and  
CC AAB76848 to AAB76878 represent human lung tumour protein related  
CC nucleotide and protein sequences which are used in the exemplification  
CC of the present invention.

12

CC NAs may be used to produce the lung-tumour associated protein, according  
 CC to standard recombinant DNA methodology. Conversely, antisense NA  
 CC molecules may be administered to down regulate protein expression by  
 CC binding with the cells own genes and preventing their expression. The NA  
 CC and complementary sequences may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar NA sequences in  
 CC samples, and hence which patients may be in need of treatment for lung  
 CC cancer. The (1) may be used as antigens in the production of antibodies  
 CC and in assays to identify modulators (agonists and antagonists) of the  
 CC expression and activity of the protein. AAF68083 to AAF68878 and  
 CC AAF68848 to AAF68878 represent human lung tumour protein related  
 CC nucleotide and protein sequences which are used in the exemplification  
 CC of the present invention.  
 XX SQ Sequence 126 AA;

Alignment Scores:  
 Pred. No.: 6.39e-47 Length: 126  
 Score: 550.00 Matches: 119  
 Percent Similarity: 52.19% Conservativity: 0  
 Best Local Similarity: 52.19% Mismatches: 1  
 Query Match: 39.86% Indels: 108  
 DB: 22 Gaps: 1

US-09-658-824-808 (1-781) x AAB76869 (1-126)

QY 2 CGCGGAGCTGTGAGCCGCGACTCGGCTCCCTGAGGTCTGGATCTTTCTCCGCTACTG 61  
 Db 7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSerProLeuLeu 26  
 QY 62 AGACACGCGGGTAGTCCACAGGCAGATCCAACTGGGAGTGAAGTGTGAGTGAGAGTG 121  
 Db 27 ArgHisGly----- 29  
 QY 122 AAGAGGAACACGACAGGCTCCGGAGGTGTGTGCTCAGTCACTCAGAGTGAGAAGGCC 181  
 Db 29 ----- 29  
 QY 182 TCGAAGTCGTGCTCCCTCTCATGGGTGTCACGCGCCATGGACCTTCTGTCTGCACGG 241  
 Db 29 ----- 29  
 QY 242 CCATAACTAGGAGGAAGAGGCGGAGAGTGGAGGGCTCAGCGCAAGCTGGGTGCT 301  
 Db 29 ----- 29  
 QY 302 GTTGGGGGTATCCAGTCCAGAGCACCTGGAACCCGACAGAGATTCGTGACTCCCC 361  
 Db 29 ----- 29  
 QY 362 AGACGGGACAGGAGGAGGCGCATGACGGACACACACACAGAACACACAGCCCA 421  
 Db 30 -----Gly--HisThrGlnThrGlnAsnHisThrAlas 40  
 QY 422 GTCCAGGAGCCAGTAATGGAGAGCCCAAAAGAAAGAACACAGCAGCTCAAGTCGCGA 481  
 Db 40 erProArgSerProValMetGluSerProLysLysLysLysLysLysLysLysLysLys 60  
 QY 482 TCCTACACCTGGCAGCAGACAGAGATCAGATACAGCTGAGATCCAGTCCGCGCA 541  
 Db 60 leuHisLeuGlySerArgGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 80  
 QY 542 CATGGAAGGTGATCTCAAGAGCTGCATCAGTCAACACCGGGGATAATCTGGATTGG 601  
 Db 80 hrTrpLysValIleCysLysSerCysLysSerGlnThrProGlyIleAsnLeuAspLeuG 100  
 QY 602 GTTCCGGCTCAAGGTGAAGATAATACCTAAAGAGGAGAACACTGTAAATGCCAGAGCAG 661  
 Db 100 lysGlyValLysValLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 120  
 QY 662 GTCAAGCAGCAACCAAGTT 681  
 Db 120 lyGluGlnProGlnVal 126

RESULT 3

AAB76866  
 ID AAB76866 Standard; Protein; 108 AA.  
 XX AAB76866;  
 AC AAB76866;  
 XX 12-APR-2001 (first entry)  
 DE Human lung tumour protein related protein sequence SEQ ID NO:786.  
 XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;  
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;  
 KW cytostatic; antisense inhibition.  
 XX Homo sapiens.  
 OS WO200100828-A2.  
 PN 04-JAN-2001.  
 PD 30-JUN-2000; 2000WO-US18061.  
 PF 30-JUN-1999; 99US-0346492.  
 XX 15-OCT-1999; 99US-0419356.  
 PR 17-DEC-1999; 99US-0466867.  
 PR 30-DEC-1999; 99US-0476300.  
 PR 06-MAR-2000; 2000US-0519642.  
 PR 22-MAR-2000; 2000US-0533077.  
 PR 10-APR-2000; 2000US-0546259.  
 PR 27-APR-2000; 2000US-0560406.  
 PR 05-JUN-2000; 2000US-0589184.  
 XX (CORI-) CORIXA CORP.  
 PA Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
 PI Retter M, Mannion J;  
 XX WPI; 2001-071488/08.  
 XX Lung tumor-associated proteins and the nucleic acids that encode them,  
 PT useful for preventing, diagnosing and treating lung cancer -  
 XX Example 1; Page 410; 436pp; English.  
 CC The present invention describes immunogenic portions of lung tumour-  
 CC associated proteins (1) and the nucleic acids (NAs) that encode them.  
 CC (1) have cytostatic activity and can be used in gene therapy, antisense  
 CC inhibition and in vaccines. The NAs and the lung tumour-associated  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with their inappropriate expression,  
 CC especially lung cancers. For example, the NAs may be administered to  
 CC treat diseases by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of the protein by expressing inactive proteins  
 CC or to supplement the patients own production of (1). Additionally, the  
 CC NAs may be used to produce the lung-tumour associated protein, according  
 CC to standard recombinant DNA methodology. Conversely, antisense NA  
 CC molecules may be administered to down regulate protein expression by  
 CC binding with the cells own genes and preventing their expression. The NA  
 CC and complementary sequences may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar NA sequences in  
 CC samples, and hence which patients may be in need of treatment for lung  
 CC cancer. The (1) may be used as antigens in the production of antibodies  
 CC and in assays to identify modulators (agonists and antagonists) of the  
 CC expression and activity of the protein. AAF68083 to AAF68878 and  
 CC AAF68848 to AAF68878 represent human lung tumour protein related  
 CC nucleotide and protein sequences which are used in the exemplification  
 CC of the present invention.  
 XX SQ Sequence 108 AA;

Alignment Scores: 3.56e-34 Length: 108  
 Pred. No.:







RESULT 7	
AA#78785	
ID	AA#78785 standard; Protein; 111 AA.
AC	AA#78785;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human protein SEQ ID NO 1447.
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
XX	nervous system disorder; arthritis; inflammation.
XX	
OS	Homo sapiens.
XX	
PN	WO200157190-A2.
XX	
PD	09-AUG-2001.
XX	
PF	05-FEB-2001; 2001WO-US04098.
XX	
PR	03-FEB-2000; 2000US-0496914.
PR	27-APR-2000; 2000US-0560875.
PR	20-JUN-2000; 2000US-0598075.
PR	19-JUL-2000; 2000US-0620325.
PR	01-SEP-2000; 2000US-0654936.
PR	15-SEP-2000; 2000US-0663561.
PR	20-OCT-2000; 2000US-0693325.
PR	30-NOV-2000; 2000US-0728422.
XX	(HYSE-) HYSEQ INC.
PA	
XX	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
PI	Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX	
DR	WPI; 2001-476283/51.
DR	N-PSDB; AAK51918.
XX	
PT	Nucleic acids encoding polypeptides with cytokine-like activities,
PT	useful in diagnosis and gene therapy -
XX	
PS	Claim 20; Page 3714-3715; 6221pp; English.
XX	
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC	encoded polypeptides (AAM78323-AA#80302) that exhibit activity elating
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation.
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC	(AAM#0020) are omitted as the relevant pages from the sequence listing
CC	were missing at the time of publication.
XX	
SQ	Sequence 111 AA;
XX	
Alignment Scores:	
Pred. No.:	6,08e-15 Length: 111
Score:	233.00 Matches: 48
Percent Similarity:	74.12% Conservative: 15
Best Local Similarity:	56.47% Mismatches: 21
Query Match:	Indels: 1
DB:	Gaps: 0
IS-09-658-824-808 (1-781) x AAM78785 (1-111)	

QY 429 GAGCCAGTAATGGAGAGCCCAAAAGAGAACAGCAGCTGAAAGTCGGGATCCTACA 488  
 Db 28 GluProThrAspGluProLysGluGluLysProProThrLysSerArgAsnProThr 47  
 QY 489 CTGCGGACGACAGACAGATCAGGATACAGCTGAGATCCAGTCGCCGACATCGAA 548  
 Db 48 ProAspGlnLysArgGluAspGlnGlyAlaAlaGluIleGlnValProAspLeuGlu 67  
 QY 549 GGTGATCTGCAAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTCGATTGGTCCGG 608  
 Db 68 AlaAspLeuGlnLysGlnThrLysThrGlyAsp-GlyCysGluGlyThrAs 87  
 QY 609 CGTCAGGTGACAGTAATCTTAACAGAGAACACTGTAATCCAGAGCAGGTGAAGA 668  
 Db 87 pValLysGlyLysIleLeuProLysAlaGluHisPheLysMetProGluAlaGlyGlu 107  
 QY 669 GCAACCCACAGTT 681  
 Db 107 yLysSerGlnVal 111  
 RESULT 8  
 AAB60500  
 ID AAB60500 standard; Protein; 111 AA.  
 AC  
 XX  
 DT 24-APR-2001 (first entry)  
 DE Human cell cycle and proliferation protein CCYPR-48, SEQ ID NO:48.  
 KW Cell cycle and proliferation protein; CCYPR; human; agonist;  
 KW antagonist; gene therapy; detection; gene therapy;  
 KW transgenic animal disease model; immune disorder;  
 KW developmental disorder; cell signalling disorder;  
 KW cell proliferative disorder; cancer; tumour; anaemia;  
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;  
 KW menstrual cycle disorder; bacterial infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200107471-A2.  
 XX  
 PD 01-FEB-2001.  
 PF 21-JUL-2000; 2000WO-US19948.  
 XX  
 PR 21-JUL-1999; 99US-0145075.  
 PR 08-SEP-1999; 99US-0153129.  
 PR 10-NOV-1999; 99US-0164647.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;  
 PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;  
 XX  
 DR WPI: 2001-112727/12.  
 XX  
 PT N-PSDB; AAF59637.  
 XX  
 PT Human cell cycle and proliferation proteins and polynucleotides are  
 used to treat, diagnose and prevent immune, developmental and cell  
 signaling disorders and cell proliferative disorders including cancer -  
 XX  
 PS Claim 1; Page 158; 205pp; English.  
 XX  
 CC Sequences AAB60453-AAB60506 represent 54 human cell cycle and  
 CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.  
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
 CC associated with decreased expression of functional CCYPR, while CCYPR  
 CC antagonists are used to treat diseases or conditions associated with  
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)

CC that specifically bind to CCYPR, and in drug screening methods to  
 CC identify compounds that modulate the activity of CCYPR. CCYPR  
 CC nucleotides can be used to generate transgenic animal models of human  
 CC disease, and can be used in gene therapy in target cells with genetic  
 CC abnormalities with respect to the expression of CCYPR for the  
 CC treatment or prevention of a disorder associated with CCYPR.  
 CC Diseases which can be diagnosed, treated and prevented using CCYPR  
 CC proteins, nucleic acids, agonists or antagonists include immune,  
 CC developmental and cell signalling disorders, and cell proliferative  
 CC disorders including cancer. Specific examples of these disorders  
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
 CC diabetes mellitus, disorders of the menstrual cycle and infections  
 XX  
 SQ Sequence 111 AA;  
 Alignment Scores:  
 Pred. No.: 6.08e-15 Length: 111  
 Score: 233.00 Matches: 48  
 Percent Similarity: 74.12% Conservative: 15  
 Best Local Similarity: 56.47% Mismatches: 21  
 Query Match: 16.88% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-658-824-808 (1-781) x AAB60500 (1-111)  
 QY 429 GAGCCAGTAATGGAGAGCCCAAAAGAGAACAGCAGCTGAAAGTCGGGATCCTACA 488  
 Db 28 GluProThrAspGluGluProLysGluGluLysProProThrLysSerArgAsnProThr 47  
 QY 489 CTGCGGACGACAGACAGATCAGGATACAGCTGAGATCCAGTCGCCGACATCGAA 548  
 Db 48 ProAspGlnLysArgGluAspGlnGlyAlaAlaGluIleGlnValProAspLeuGlu 67  
 QY 549 GGTGATCTGCAAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTCGATTGGTCCGG 608  
 Db 68 AlaAspLeuGlnLysGlnThrLysThrGlyAsp-GlyCysGluGlyThrAs 87  
 QY 609 CGTCAGGTGACAGTAATCTTAACAGAGAACACTGTAATCCAGAGCAGGTGAAGA 668  
 Db 87 pValLysGlyLysIleLeuProLysAlaGluHisPheLysMetProGluAlaGlyGlu 107  
 QY 669 GCAACCCACAGTT 681  
 Db 107 yLysSerGlnVal 111  
 RESULT 9  
 AAB60500  
 ID AAB60500 standard; Protein; 249 AA.  
 AC  
 XX  
 DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #27039.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.  
 DR N-PSDB; AAS91235.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID NO 57407; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 249 AA;

Alignment Scores:  
 Pred. No.: 5.84e-14 Length: 249  
 Score: 224.50 Matches: 55  
 Percent Similarity: 60.19% Conservative: 10  
 Best Local Similarity: 50.93% Mismatches: 33  
 Query Match: 16.27% Indels: 10  
 DB: 22 Gaps: 1

US-09-658-824-808 (1-781) x ABG27048 (1-249)

QY 399 ACAACACAGACACACAGCCAGTCCAGGAGCCCGATGAGAGCCCAAAAGAA 458  
 Db 61 ThrSerPheLeuProThrMetThrGlnGlnGluProSerValProGluProGlnGlu 80  
 QY 459 GAACACAGCAGCTGAAAGTCGGGATCTACACCTGGCGACGACAGACAGATCAGGAT 518  
 Db 81 GluProProThrGluSerGlnAspHisThrProGlyGlnLysArgGluAspGlnGly 100  
 QY 519 ACAGCTGAGATCCAGTGGCGGACATGGAGGTGATCTCAAGAGCTGCATCAGTCAAC 578  
 Db 101 AlaAlaGluIleGlnValProAsnLeuGluAlaAspLeuGlnGluSerGlnSerLys 120  
 QY 579 ACCGGGATAAATCTGGATTGGGTTCGCGCGTCAAGGTGAAGATAATACCTAAACAGGA 638  
 Db 121 ThrGlyAspGluCysGlyAspSerProAspValGlnGly-LysIleLeuProLysSerG 140  
 QY 639 ACATGTAAATGCCAGAA-----GCAGGTGAAGAGCA 671  
 Db 140 uGlnPheLysMetProGluGlyGluHisLeuArgAspValIleHisAlaCysGluGluG 160  
 QY 672 ACCAAGATTAAATGAAGACA 693  
 Db 160 uGluProValSerMetThrThr 167

RESULT 10  
 AAM41374  
 ID AAM41374 standard; Protein; 106 AA.  
 XX

AAM41374;  
 22-OCT-2001 (first entry)  
 Human polypeptide SEQ ID NO 6305.  
 Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 peripheral nervous system; neuropathy; central nervous system; CNS;  
 Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;  
 amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 leukaemia.  
 XX Homo sapiens.  
 OS WO200153312-A1.  
 PN 26-JUL-2001.  
 PD 26-DEC-2000; 2000WO-US34263.  
 PF 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 Wang J, Jang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 WPI: 2001-442253/47.  
 N-PSDR; AAI60530.  
 Novel nucleic acids and polypeptides, useful for treating disorders  
 such as central nervous system injuries -  
 Example 2; SEQ ID NO 6305; 10078pp; English.  
 The invention relates to human nucleic acids (AA157798-AA161369) and  
 the encoded polypeptides (AAM38642-AA42213) with nootropic,  
 immunosuppressant and cytostatic activity. The polynucleotides are useful  
 in gene therapy. A composition containing a polypeptide or polynucleotide  
 of the invention may be used to treat diseases of the peripheral nervous  
 system, such as peripheral nervous injuries, peripheral neuropathy and  
 localized neuropathies and central nervous system diseases, such as  
 Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 utilisation of the activities such as: immune system suppression,  
 Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 assays for receptor activity, arthritis and inflammation, leukaemias and  
 C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 SQ Sequence 106 AA;  
 Alignment Scores:  
 Pred. No.: 1.23e-13 Length: 106  
 Score: 220.00 Matches: 47  
 Percent Similarity: 68.83% Conservative: 6  
 Best Local Similarity: 61.04% Mismatches: 23  
 Query Match: 15.94% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-658-824-808 (1-781) x AAM41374 (1-106)



DT	24-JUL-2000	(first entry)	
XX	PAGE-4	polypeptide.	
XX	PAGE-4	MAGE; GAGE; reproduction; testis; prostate; fallopian tube;	
KW	uterus; placenta; cancer; major histocompatibility complex; MHC; CTL		
KW	cytotoxic T lymphocyte; immune response; antibody; drug delivery;		
KW	immunconjugate.		
XX	Homo sapiens.		
OS	WO200012706-A1.		
PN	09-MAR-2000.		
XX	31-AUG-1999;	99WO-US20046.	
XX	01-SEP-1998;	98US-0098993.	
XX	(USSH ) US DEPT HEALTH & HUMAN SERVICES.		
PA	Pastan I, Brinkmann U, Vasmatazis G, Lee B;		
PI	WPI; 2000-237869/20.		
XX	N-PSDB; AAZ93524.		
DR	Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T		
PT	lymphocyte response and for raising antibodies which can be used to		
PT	detect the presence of PAGE-4 in cell samples or body tissues		
XX	Disclosure; Figure 1; 63pp; English.		
PS	PAGE-4 is a gene preferentially expressed in normal male and female		
XX	reproductive tissues e.g. prostate, testis, fallopian tube, uterus		
CC	and placenta, as well as in prostate cancer, testicular cancer and		
CC	uterine cancer. This expression pattern makes it a target for		
CC	diagnosis and for vaccine based therapy of such neoplasms.		
CC	An isolated PAGE-4 peptide which induces a cytotoxic T		
CC	lymphocyte response when bound to a major histocompatibility complex		
CC	(MHC) class I molecule or the isolated PAGE-4 protein can be used in		
CC	immunogenic compositions to raise a cytotoxic T lymphocyte response		
CC	against cells expressing PAGE-4 including cancer cells of the		
CC	prostate, uterus and testis. The nucleic acids encoding PAGE-4 or		
CC	PAGE-4 peptide fragments can also be used in these compositions.		
CC	Antibodies against PAGE-4 and its peptide fragments can be used in		
CC	detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell		
CC	samples or body tissues. The presence of PAGE-4 in tissues which are		
CC	not related to reproduction can be indicative of the spread of		
CC	cancerous reproductive tissue. PAGE-4 can also be used to raise		
CC	antibodies which are then used as the targeting group of		
CC	immunconjugates comprising toxins used in therapeutic applications.		
CC	This has applications for drug delivery systems.		
XX	Sequence	102 AA;	
SQ	Alignment Scores:		
	Pred. No.:	2,04e-05	Length: 102
	Score:	138.50	Matches: 34
	Percent Similarity:	55.70%	Conservative: 10
	Best Local Similarity:	43.04%	Mismatches: 31
	Query Match:	10.04%	Indels: 4
	DB:	21	Gaps: 1
US	-09-658-824-808 (1-781) x AAY83158 (1-102)		
QY	441 GGAGAGCCCAAAAGAAACAGCAGCTGAAGTCGGATCCTACACTGGCAGCAG	500	
Db	28 GlycluserGlnGluGluProThrPasnGlnAspIleGluProGlyGlnGlu	47	
QY	501 ACAGAAGATCAGGATACAGCTGAGATCCCAAGTCGGAGATGAAGGTGATCTGCAA	560	
Db	48 ArgGlu-----GlyThrProProIleGluGluArgLvsValGluGlvAspCvsGln	64	

QY 561 GAGTCGATCAGTCAAAACACCGGGGATAAAATCTGGATTTCGGTTCCGGCGTCAAGGTGAA 620  
 Db 65 GluMetAspLeuGluLysThrArgSerGluArgGlyAsp-GlySerAspValLysGluLys 84  
 QY 621 GATAATACCTAAAGAGGAACACCTGTAATATGCCAGAGCAGGTGACAGAGCAACA 675  
 Db 84 sThrProAsnProLysHisAlaLysThrLysGluAlaGlyAspGlyGlnPro 102

RESULT 14  
 AAY52998  
 ID AAY52998 standard; Protein; 102 AA.  
 AC AAY52998;  
 XX 29-FEB-2000 (first entry)  
 XX Human secreted protein clone bn365\_53 protein sequence SEQ ID NO:2.  
 DE Human; secreted protein; nutritional; cytokine; cell proliferation;  
 KW differentiation; immune stimulating; vaccine; suppression;  
 KW haematopoiesis regulation; tissue growth; activin; inhibin;  
 KW chemotactic; chemokinetic; haemostatic; thrombolytic; receptor;  
 KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;  
 KW tumour inhibition; gene therapy.  
 XX Homo sapiens.  
 OS  
 XX  
 PN W09957132-A1.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 07-MAY-1999; 99WO-US09970.  
 XX  
 PR 07-MAY-1998; 98US-0084564.  
 PR 02-JUN-1998; 98US-0087645.  
 PR 22-JUL-1998; 98US-0093712.  
 PR 31-JUL-1998; 98US-0094935.  
 PR 10-AUG-1998; 98US-0095880.  
 PR 11-AUG-1998; 98US-0096068.  
 PR 06-MAY-1999; 99US-0096068.  
 XX (GENY ) GENETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Bowman MR;  
 PI DiBlasio-Smith E, Widom A;  
 XX  
 DR WPI; 2000-052937/04.  
 DR N-PSDB; AAZ33316.  
 XX  
 PT New polynucleotides encoding secreted human proteins, derived from  
 adult placenta, adult retina, fetal brain, fetal  
 XX  
 PS Claim 9; Page 360-361; 492pp; English.  
 XX  
 CC The present invention describes new human secreted proteins which were  
 isolated from adult placenta, adult retina, foetal brain, foetal kidney,  
 CC adult blood, adult brain, adult thyroid, adult bladder, adult neural  
 CC tissue, adult testes, and adult lymph node cDNA libraries. The human  
 CC secreted proteins, and the polynucleotides encoding them, are predicted  
 CC to have biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals. Suggested activities include nutritional activity, cytokine  
 CC and cell proliferation/differentiation activity, immune stimulating  
 CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating  
 CC activity, tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
 CC activity, receptor/ligand activity, anti-inflammatory activity,  
 CC cadherin/tumour invasion suppressor activity, and tumour inhibition  
 CC activity. The polynucleotides are also stated to be useful for gene  
 CC therapy. AAZ33316 to AAZ33373 encode human secreted proteins, and  
 CC AAY52998 to AAY53060 represent human secreted proteins, given in the  
 CC present invention.

XX Sequence 102 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 2,04e-05 Length: 102  
 Score: 138.50 Matches: 34  
 Percent Similarity: 55.70% Conservative: 10  
 Best Local Similarity: 43.04% Mismatches: 31  
 Query Match: 10.04% Indels: 4  
 DB: 21 Gaps: 1

US-09-658-824-808 (1-781) x AAY52998 (1-102)  
 QY 441 GGAGAGCCCCAAAAGAACACAGCAGCTGAAGTCTGGATCTACACTGGCGACAG 500  
 Db 28 GlyGluSerGlnGlnGluProThrAspAsnGlnAspIleGluProGlyGlnGlu 47  
 QY 501 ACAGAGAAGATCAGGATCAGCTGAGATCCAGTCCGACATGAAGTGTATCTGCAA 560  
 Db 48 ArgGlu-----GlyThrProIleGluGluArgLysValGluGlyAspCysGln 64  
 QY 561 GAGTCGATCAGTCAAAACACCGGGGATAAATCTGGATTTCGGCGTCAAGGTGAA 620  
 Db 65 GluMetAspLeuGluLysThrArgSerGluArgGlyAsp-GlySerAspValLysGluLys 84  
 QY 621 GATAATACCTAAAGAGGAACACCTGTAATATGCCAGAGCAGGTGACAGAGCAACA 675  
 Db 84 sThrProAsnProLysHisAlaLysThrLysGluAlaGlyAspGlyGlnPro 102

RESULT 15  
 AAY12447  
 ID AAY12447 standard; Protein; 89 AA.  
 AC AAY12447;  
 XX 17-JUN-1999 (first entry)  
 XX Human 5' EST secreted protein SEQ ID NO:478.  
 DE  
 XX Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09906548-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB01222.  
 XX  
 PR 01-AUG-1997; 97US-0905135.  
 XX (GENST ) GENSET.  
 XX  
 PA Duclert A, Dumas Milne Edwards J, Lacroix B;  
 PI  
 XX WPI; 1999-153778/13.  
 DR N-PSDB; AAX41280.  
 XX  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
 PT kidney, lung, umbilical cord, placenta and colon tissue  
 XX  
 PS Claim 27; Page 778; 824pp; English.  
 XX  
 CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for  
 human secreted proteins, and encode the proteins given in AAY12261 to  
 CC AAY12514, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid









Db	341	euSerProGlnProAlaSerProAlaGluClnuAspLysMetProProTyrAspGluG	361
Qy	499	AGACAGCAAGATCAGGATACA-----GCTGAGA	528
Db	361	lnThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysPheGluAlaGluA	381
Qy	529	TCCCAAGTCGCGACATGGAAGGTGCTGCAAGAGCTGCATCAGTCAACACCGGGGATA	588
Db	381	rgSerLeuTysAspMetGluGluSerIleArgAsnLeuGluGln-----G	396
Qy	589	AATCTGATGTGGTTCGCGCGTCAAGGTGAA	620
Db	396	luIleSerPheAspPheGlyProAsnGlyGlu	406
RESULT	20		
AAU28042	ID	AAU28042 standard; Protein: 527 AA.	
XX	AC	AAU28042;	
XX	AC		
XX	DT	18-DEC-2001 (first entry)	
XX	XX	Novel human secretory protein, Seq ID No 211.	
XX	XX	Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;	
KW	KW	ischemia-reperfusion injury; haematopoiesis; cancer; neuropathy;	
KW	KW	transgenic animal; Alzheimer's disease; Parkinson's disease; burn;	
KW	KW	autotrophic lateral sclerosis; platelet disorder; thrombocytopenia;	
KW	KW	ulcer; osteoporosis; bone degenerative disorder; periodontal disease;	
KW	KW	gut protection; lung; liver fibrosis; immune deficiency; infection;	
KW	KW	severe combined immunodeficiency; SCID; autoimmune disorder; allergy;	
KW	KW	multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;	
KW	KW	fertility; analgesic; pain; antigen.	
XX	XX		
OS	OS	Homo sapiens.	
XX	XX		
PN	PN	WO200166689-A2.	
XX	XX	13-SEP-2001.	
XX	XX		
XX	XX	05-MAR-2001; 2001WO-US04942.	
XX	XX		
PR	PR	07-MAR-2000; 2000US-0519705.	
PR	PR	19-MAY-2000; 2000US-0574454.	
PR	PR	17-JUN-2000; 2000US-0596193.	
PR	PR	14-JUL-2000; 2000US-0616847.	
PR	PR	19-SEP-2000; 2000US-0665363.	
PR	PR	20-OCT-2000; 2000US-0693267.	
XX	XX		
PA	PA	(HYSE-) HYSEQ INC.	
XX	XX		
PI	PI	Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P	
PI	PI	Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;	
XX	XX		
DR	DR	WPI: 2001-589934/66.	
DR	DR	N-PSDB: AAS44942.	
XX	XX		
PT	PT	Novel polypeptides and nucleic acids obtained from cDNA libraries	
PT	PT	prepared from various human tissues, for diagnosis and treatment of	
PT	PT	cancer, neurological, inflammatory, and autoimmune disorders -	
PS	PS	Example 3; SEQ ID No 211; 107pp; English.	
XX	XX		
CC	CC	The invention relates to novel isolated human secreted polypeptides (	
CC	CC	and polynucleotides (II). (I) and (II) are useful for treating	
CC	CC	inflammatory conditions such as arthritis, nephritis, Crohn's disease	
CC	CC	ischemia-reperfusion injury, shock, sepsis, immune responses, and is	
CC	CC	involved in increasing haematopoiesis, stem cell survival, bone growth	
CC	CC	and remodeling. (I), (II) and modulators of (II) are useful for	
CC	CC	prophylaxis or treatment of one or more cancers. (II) is also useful	
CC	CC	creating transgenic animals useful for studying the in vivo activities	
CC	CC	the polypeptide as well as for studying modulators of the polypeptides	
CC	CC	(I) induces the proliferation of neural cells and regeneration of nerve	



PA (UYCO ) UNIV COLUMBIA NEW YORK.  
XX Fisher PB, Goldstein NI, Su Z, Zhang N;  
XX WPI; 1998-557025/47.  
DR N-PSDB; AAV65765.  
XX New isolated progression Elevated Gene-3 - used to develop products  
PT for e.g. modulating DNA damage and repair pathways, cancer  
PT progression or oncogene mediated transformation and angiogenesis.  
XX Claim 17; Fig 13A-C; 225pp; English.  
XX This polypeptide comprises human progression elevated gene-3 (PEG-3)  
CC protein. The amino acid sequence was deduced from an isolated cDNA  
CC clone (see AAV65765). PEG-3 expression (1.5 and 2.8 kb transcripts)  
CC is elevated in most human tumour cell lines. The amount of PEG-3  
CC in a cell is also an indicator of DNA damage in that cell. Cells  
CC in which a reporter gene is under control of the promoter or  
CC regulatory region (see AAV65766) of the rat PEG-3 gene can be used in  
CC methods for identifying agents that modulate PEG-3 expression or the  
CC ability of PEG-3 to induce progression, or for determining whether  
CC an agent is capable of inhibiting DNA damage and repair pathways,  
CC cancer progression or oncogene-mediated transformation. Compounds  
CC that induce DNA damage or which regulate angiogenesis can also be  
CC identified using such cells. Transgenic animals and vaccines  
CC comprising PEG-3 polypeptides and an immune response enhancer are  
CC also claimed. Antibodies (especially monoclonal) to the PEG-3 may  
CC be used to determine tumour progression of a cell. Cells can be  
CC protected from chemotherapeutic damage by inhibiting or eliminating  
CC the expression of PEG-3 in the cells. The methods can be applied  
CC to a progression phenotype comprising anchorage-independent growth,  
CC tumorigenesis, angiogenesis or metastasis, to melanoma, brain,  
CC cervical, prostate, lung or colorectal cancer, neuroblastoma or  
CC glioblastoma (all claimed).  
XX Sequence 578 AA;

Alignment Scores:  
Pred. No.: 0.145 Length: 578  
Score: 103.00 Matches: 49  
Percent Similarity: 37.99% Conservative: 19  
Best Local Similarity: 27.37% Mismatches: 67  
Query Match: 7.46% Indels: 44  
Gaps: 9

US-09-658-824-808 (1-781) x AAV79958 (1-578)

QY 112 AGTGTAGAGTGAAGAGAACACAGCAGGCTTCCGGAGGGTTGTGTGCTGACTCAGAGT 171  
DB 220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgPheTrpGluTyrHisThrArg 239  
QY 172 GAGAAGCCCTCGAAGTCGTGCTCCCTCATCGGTGCGCCAGCCCATGGACCTTCTGT 231  
DB 240 GluArgProLysGlnGluGlyThrLysProGluGlnHisArgAlaGlyGlnSerHis 259  
QY 232 CTCGTACGCCATACTAGGAGGAGGAGGCCCGCAGG-----AGTGGAGG 279  
DB 260 ProCysGlnAsnAlaGluSerGluGluGlyProGluThrSerValCysSerGlySer 279  
QY 280 GCTCAGCGAAGCTGGGTGCTGTGGGGGTATCCGAGTCCAGAACACCTGGAAACCCC 339  
DB 280 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 292  
QY 340 GACAGAAATCTCGACTCCCA-----GACGGGACAGGAGGAGCGGATGAGCGA 393  
DB 292 pThrGluGluGlu-GluAspProAspLeuAspSerAlaGluGluAspThrAla----- 309  
QY 394 CACACAAACACACAGAACACACAGC-----  
DB 310 --HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 329  
QY 420 -----CAGTCCAGGAGGCCAGTAATGGAGAG---CCCAAAAGAGAACACAGCAGCTG 471

DB 329 lyGluAspThrGluGluGluAspGlyAspTrpAspSerAlaGluGluAspAlaAlaG 349  
QY 472 AAAGf---CGGATCCT-----ACACCTGGGC 495  
DB 349 InSerCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 369  
QY 496 AGCAGACAGAGAGATCAGATACAGCTGAGATCCAGTCCGCGGACATGAA 548  
DB 369 luAspThrGluGluGluAspSerGluAsnValAlaProValAspSerGlu 386

RESULT 23

AAAY41104  
ID AAY41104 standard; Protein: 578 AA.  
XX  
AC AAY41104;  
XX  
DT 17-JAN-2000 (first entry)  
XX  
DE Human progression elevated gene-3 (PEG-3) amino acid sequence.  
XX  
KW Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;  
KW cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; human.  
XX  
OS Homo sapiens.  
XX  
PN MO9949898-A1.  
XX  
PD 07-OCT-1999.  
XX  
PF 31-MAR-1999; 99WO-US07199.  
XX  
PR 31-MAR-1998; 98US-0052753.  
XX  
PA (UYCO ) UNIV COLUMBIA NEW YORK.  
XX  
PI Fisher PB;  
XX  
DR WPI; 1999-591184/50.  
XX  
N-PSDB; AAZ23029.  
XX  
PT Novel vectors useful for studying the progression of cancer -  
PS  
Disclosure; Fig 13A-C; 251pp; English.  
XX  
CC The invention relates to an inducible progression-elevated gene-3 (PEG-3  
CC gene) regulatory region functionally linked to a gene encoding a product  
CC that causes or may be induced to cause the death or inhibition of cancer  
CC cell growth. A vector of the invention which contains a gene encoding a  
CC thymidine kinase or a product which causes the cell to express a  
CC specific antigen can be administered along with gancyclovir or acyclovir,  
CC or an antibody or fragment to the antigen, respectively, to treat cancer  
CC in a subject. The PEG-3 gene is useful for generating new cloning and  
CC expression vectors, transfected cells, and for developing methods for  
CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as  
CC a source of primers and probes to study the progression of cancer, and to  
CC detect the presence of the gene. The present sequence represents the  
CC amino acid sequence of the human PEG-3 protein.  
XX  
SQ Sequence 578 AA;

Alignment Scores:  
Pred. No.: 0.145 Length: 578  
Score: 103.00 Matches: 49  
Percent Similarity: 37.99% Conservative: 19  
Best Local Similarity: 27.37% Mismatches: 67  
Query Match: 7.46% Indels: 44  
Gaps: 9

US-09-658-824-808 (1-781) x AAY41104 (1-578)

QY 112 AGTGTAGAGTGAAGAGAACACAGCAGGCTTCCGGAGGGTTGTGTGCTGACTCAGAGT 171  
DB 112

Db 220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgPheTrpGluTrpHisThrArg 239  
 QY 172 GAGAAGCCCTCGAAGTCGTCCTCTCATCGGTGCACGCCCATGACCTTCTTGT 231  
 Db 240 GluArgProLysGlnGluGlyGluThrLysProGluGlnHisArgAlaGlyGlnSerHis 259  
 QY 232 CTCGTACGCCCATAACTAGGAGGAGGAGGCCGAGG-----AGTGGAGG 279  
 Db 260 ProCysGlnAsnAlaGluSerGluGlyGlyProGluThrSerValCysSerGlySer 279  
 QY 280 GCTCAGCGGAAGCTGGGGTCTGTTGGGGTATCCGAGTCCCGAGACCTGGAACCCC 339  
 Db 280 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 292  
 QY 340 GACAGAGATTCGTGGACTCCCA-----CACGGGACAGAGAGCGGCATGACCGA 393  
 Db 292 pThrGluGluGlu-GluAspProAspLeuAspSerAlaGluGluAspThrAla----- 309  
 QY 394 CACACACAAACACAGAACCCACACAGC----- 419  
 Db 310 --HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 329  
 QY 420 -----CAGTCCCGAGGCCAGTAATGGAGAG---CCCCAAAAGAGAACACGACGTG 471  
 Db 329 lyGluAspThrGluGluGluAspGlyAspTrpAspSerAlaGluGluAspAlaAlaG 349  
 QY 472 AAAGT---CGGGATCCT----- 495  
 Db 349 InSerCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 369  
 QY 496 AGCAGACAGAAAGATCAGGATACAGTCCAGTCCGCGGACATGGAA 548  
 Db 369 luAspThrGluGluGluAspSerGluAsnValAlaProValAspSerGlu 386  
 RESULT 24  
 AAY06514  
 ID AAY06514 standard; Protein; 578 AA.  
 XX  
 AC AAY06514;  
 XX  
 DT 08-OCT-1999 (first entry)  
 XX  
 DE Human progression elevated protein-3 (PEG-3).  
 XX  
 KW Progression elevated protein-3; PEG-3; human; angiogenesis; cancer;  
 KW therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9937776-A1.  
 XX  
 PD 29-JUL-1999.  
 XX  
 PF 26-JAN-1999; 99WO-US01623.  
 XX  
 PR 29-JAN-1998; 98US-0072941.  
 PR 26-JAN-1998; 98US-0072469.  
 XX  
 PA (GENQ-) GENQUEST INC.  
 XX  
 PI Fisher PB, Zhang N;  
 XX  
 DR WPI; 1999-458694/38.  
 DR N-PSDB; AAX87383.  
 XX  
 PT Modulation of angiogenesis by altering the expression and/or  
 PT activity of a progression-associated protein, especially for cancer  
 PT treatment  
 XX  
 PS Claim 1; Page 74-76; 81pp; English.  
 PS  
 CC The present sequence represents human progression elevated protein-3  
 CC (PEG-3), as deduced from a cDNA clone (see AAX87383) isolated from a

CC MCF-7 cDNA library. PEG-3 represents a novel member of the  
 CC gadd34/MyD116 gene family. Human PEG-3 is generally expressed in  
 CC cells that are in progression, including most tumour cell lines. A  
 CC claimed method for modulating angiogenesis in an organism comprises  
 CC administering an agent that alters expression and/or activity of a  
 CC PEG-3 protein, such as an antisense polynucleotide or antibody.  
 CC Also claimed is a method for determining whether an agent  
 CC modulates angiogenesis, where the candidate agent is present  
 CC within a combinatorial small molecule library. Vaccines and  
 CC pharmaceutical compositions comprising such compounds are also  
 CC provided and may be used to prevent angiogenesis, especially  
 CC related to cancer cell progression.  
 XX  
 SQ Sequence 578 AA;  
 Alignment Scores:  
 Pred. No.: 0.145 Length: 578  
 Score: 103.00 Matches: 49  
 Percent Similarity: 37.99% Conservative: 19  
 Best Local Similarity: 27.37% Mismatches: 67  
 Query Match: 7.46% Indels: 44  
 DB: 20 Gaps: 9  
 US-09-658-824-808 (1-781) x AAY06514 (1-578)  
 QY 112 AGTGAGAGTGAAGAGAACACGAGCGCTTCCGAGGGTTGTGGTCAGTCACTCAGAGT 171  
 Db 220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgPheTrpGluTrpHisThrArg 239  
 QY 172 GAGAAGCCCTCGAAGTCGTCCTCTCATCGGTGCACGCCCATGACCTTCTTGT 231  
 Db 240 GluArgProLysGlnGluGlyGluThrLysProGluGlnHisArgAlaGlyGlnSerHis 259  
 QY 232 CTCGTACGCCCATAACTAGGAGGAGGAGGCCGCGAGG-----ACTGGAGGG 279  
 Db 260 ProCysGlnAsnAlaGluSerGluGlyGlyProGluThrSerValCysSerGlySer 279  
 QY 280 GCTCAGCGGAAGCTGGGGTCTGTTGGGGTATCCGAGTCCCGAGACCTGGAACCCC 339  
 Db 280 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 292  
 QY 340 GACAGAGATTCGTGGACTCCCA-----CACGGGACAGGAGGCGGCATGACCGA 393  
 Db 292 pThrGluGluGlu-GluAspProAspLeuAspSerAlaGluGluAspThrAla----- 309  
 QY 394 CACACACAAACACAGAACCCACACAGC----- 419  
 Db 310 --HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 329  
 QY 420 -----CAGTCCCGAGGCCAGTAATGGAGAG---CCCCAAAAGAGAACACGACGTG 471  
 Db 329 lyGluAspThrGluGluGluAspGlyAspTrpAspSerAlaGluGluAspAlaAlaG 349  
 QY 472 AAAGT---CGGGATCCT----- 495  
 Db 349 InSerCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 369  
 QY 496 AGCAGACAGAAAGATCAGGATACAGTCCAGTCCGCGGACATGGAA 548  
 Db 369 luAspThrGluGluGluAspSerGluAsnValAlaProValAspSerGlu 386  
 RESULT 25  
 ID AAB95541  
 XX  
 AC AAB95541;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:18149.  
 XX  
 KW Human; ; imer; aetection; diagnosis; antisense therapy; gene therapy.  
 KW











KW transcription activation; DEAD box RNA dependent helicase;  
KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.  
XX  
OS Homo sapiens.  
PN WO200073467-A1.  
XX  
PD 07-DEC-2000.  
XX  
PF 25-MAY-2000; 2000WO-US14719.  
XX  
PR 27-MAY-1999; 99US-0136620.  
PR 25-MAY-2000; 2000US-0579181.  
XX  
PA (UYSL-) UNIV SAINT LOUIS.  
XX  
PI Chirivia J, Yaciuk P;  
XX  
DR WPI; 2001-061545/07.  
DR N-PSDB; AAC89860.  
XX  
PT Snf2 related cAMP regulatory element (CREB) binding protein (CBP)  
PT activator protein, capable of co-activating CREB binding protein,  
PT useful for modulating transcription and for affecting viral infection -  
XX  
PS Claim 1; Page 86-94; 103pp; English.  
XX  
CC The present sequence is an Snf2 related CREB (cAMP regulatory element)  
CC binding protein (CBP) activator protein (SRCAP) polypeptide. It has  
CC ATPase activity and is capable of activating transcription. SRCAP  
CC polypeptides are useful for activating transcription in a cell, for  
CC enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated  
CC activation of transcription in a cell, for treating a patient having a  
CC disease involving a function such as insufficient transcription of a  
CC gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent  
CC helicase, adenoviral DBP protein, beta-actin or a nuclear receptor  
CC affected by SRCAP protein. Compounds that modulate SRCAP function, such  
CC as antibodies, antisense molecules, polynucleotides or ribozymes, are  
CC useful for treating diseases mediated by SRCAP-activated transcription,  
CC for example, infection by adenovirus, hepatitis C virus, human  
CC immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or  
CC hepatitis B virus.  
XX  
SQ Sequence 2972 AA:  
  
Alignment Scores:  
Pred. No.: 0.291 Length: 2972  
Score: 102.50 Matches: 70  
Percent Similarity: 42.11% Conservative: 34  
Best Local Similarity: 28.34% Mismatches: 81  
Query Match: 7.19% Indels: 62  
DB: 22 Gaps: 12  
  
US-09-658-824-808 (1-781) x AAB50363 (1-2972)  
QY 640 GTTCCTTTTAGTATTATCTTCCACCTTGACGCGGAAACCCAAATGCAGAT----- 590  
Db 1291 ValProAlaSerAlaLeuAlaSerProPheProSerAlaProAlaProAlaGln 1310  
QY 589 -----TTATCCCGGTGTTGACTGATGACGCTCTTGACATCACCTTCCATGTGC 539  
Db 1311 AlaSerLeuAlaPro-----AlaSerSerAlaSerGlnAlaLeuAla 1325  
QY 538 CGCACTGGGATCTCAGCTGATCTGATCTCTCTGCTGCTGCTGCCAGGTGTAGATCC 479  
Db 1326 ThrProLeuAlaProMetaAlaAlaProGlnThrAlaLeuAlaProSerProAlaPro 1345  
QY 478 CGACTTTCAGCTGCTGCTTCTCTTTTGGGGCTCTCCATTTACTGGGCTCTGGGACTGG 419  
Db 1346 ProLeuAlaProLeu-----ProValLeu-AlaProSerProGl 1358  
QY 418 CTGTGTG---GTTCTGTGTTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362

Db 1358 yAlaAlaProValLeuAlaSerSerGlnThrProValProValMet----- 1373  
QY 361 GGGAGTCCAGAAATCTTCTCGGGGTCCAGGTGCTTCTGGGACTCGGATACCCCAAC 302  
Db 1374 ----AlaProSerThrProGlyThrSerLeuAlaSerAlaSerProValPro----- 1390  
QY 301 AGCACCCAGCTTCGCTGAGCCCTCCACT-----CCTCGGCC 263  
Db 1391 -AlaProThrProValLeuAlaProSerSerThrGlnThrMetLeuProAlaProValPr 1410  
QY 262 CTCCTTCTCCTAGTTATGGCGGTGACGAGACAGAGAGAGGTCCATGGCGTGGCAGCGCA 203  
Db 1410 oSerProLeuProSerProAlaSerThrGlnThr-----LeuAlaLeuAlaPro-A 1427  
QY 202 TGAGAGGAGCAGCAGCTTCGAGGGCCTCTCAGCTCTCAGTCACTG-----ACCACA 152  
Db 1427 laLeuAlaProThr-LeuGlyGlySerSerProSerGlnThrLeuSerLeuGlyThrGly 1446  
QY 151 CAACCTCCGGAAGCCTGCTGCTTCTTCTCCTCTTCACTCTCCTC----- 111  
Db 1447 AsnProGlnGly--ProPheProThrGlnThrLeuSerLeuThrProAlaSerSerLeuV 1466  
QY 110 -----ACACTTCAACTCCAGTTGGATCTGCTGTGGAGCTACCGGCC 68  
Db 1466 alProThrProAlaGlnThrLeuSerLeuAlaProGlyProProLeuGlyProThrGlnT 1486  
QY 67 GTGTCTCAGTAGCGGAG-----AAAGAAATCCAGACCTCAGGGAGCCGAGTGC 20  
Db 1486 hrLeuSerLeuAlaProAlaProProLeuAlaProAlaSerProValGlyPro---AlaP 1505  
QY 19 CGGCTCACAGC 9  
Db 1505 roAlaHisThr 1508  
RESULT 31  
AAB50362  
ID AAB50362 standard; protein; 3118 AA..  
XX  
AC AAB50362;  
XX  
DT 12-MAR-2001 (first entry)  
XX  
DE Human SRCAP.  
XX  
KW Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;  
KW CAMP regulatory element; CREB binding protein; CBP; ATPase;  
KW transcription activation; DEAD box RNA dependent helicase;  
KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.  
OS Homo sapiens.  
XX  
PN WO200073467-A1.  
XX  
PD 07-DEC-2000.  
XX  
PF 25-MAY-2000; 2000WO-US14719.  
XX  
PR 27-MAY-1999; 99US-0136620.  
PR 25-MAY-2000; 2000US-0579181.  
XX  
PA (UYSL-) UNIV SAINT LOUIS.  
XX  
PI Chirivia J, Yaciuk P;  
XX  
DR WPI; 2001-061545/07.  
DR N-PSDB; AAC89859.  
XX  
PT Snf2 related cAMP regulatory element (CREB) binding protein (CBP)  
PT activator protein, capable of co-activating CREB binding protein,  
PT useful for modulating transcription and for affecting viral infection -  
XX  
PS Claim 6; Page 77-86; 103pp; English.

CC The present sequence is an Snf2 related CREB (cAMP regulatory element)  
 CC binding protein (CBP) activator protein (SRAP) polypeptide. It has  
 CC ATPase activity and is capable of activating transcription. SRAP  
 CC polypeptides are useful for activating transcription in a cell, for  
 CC enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated  
 CC activation of transcription in a cell, for treating a patient having a  
 CC disease involving a function such as insufficient transcription of a  
 CC gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent  
 CC helicase, adenoviral DBP protein, beta-actin or a nuclear receptor  
 CC affected by SRAP protein. Compounds that modulate SRAP function, such  
 CC as antibodies, antisense molecules, polynucleotides or ribozymes, are  
 CC useful for treating diseases mediated by SRAP-activated transcription,  
 CC for example, infection by adenovirus, hepatitis C virus, human  
 CC immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or  
 CC hepatitis B virus.  
 XX  
 CC  
 SQ Sequence 3118 AA;

Alignment Scores:  
 Pred. No.: 0.296 Length: 3118  
 Score: 102.50 Matches: 70  
 Percent Similarity: 42.11% Conservative: 34  
 Best Local Similarity: 28.34% Mismatches: 81  
 Query Match: 7.19% Indels: 62  
 DB: 22 Gaps: 12

US-09-658-824-808 (1-781) x AAB50362 (1-3118)

QY 640 GTTCTCTTTAGTATATCTTCACCTTGACCGCGAACCCAAATCCAGAT----- 590  
 DB 1437 ValProAlaSerAlaLeuAlaSerProPheProSerAlaProAsnProAlaProAlaGln 1456  
 QY 589 -----TTATCCCGGTGTTTCTGACTGATGACGCTCTTGCGATGACCTTCCATGTCG 539  
 DB 1457 AlaSerLeuLeuAlaPro-----AlaSerSerAlaSerGlnAlaLeuAla 1471  
 QY 538 CGCACTGGGATCTACGCTGATCTTCTCTGCTGCTGCGCCAGGTAGGATCC 479  
 DB 1472 ThrProLeuAlaProMetAlaAlaProGlnThrAlaLeuAlaProSerProAlaPro 1491  
 QY 478 CGACTTTCAGCTGCTGGTCTCTCTTTGGGCTCTCCATTAAGTCTGGGCTGGGACATGG 419  
 DB 1492 ProLeuAlaProLeu-----ProValLeu-AlaProSerProGln 1504  
 QY 418 CTGTGTG---GTCTCTGTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362  
 DB 1504 YAlaAlaProValLeuAlaSerSerGlnThrProValProValMet----- 1519  
 QY 361 GGGAGTCCAGATCTTCTGCTGGGTTCCAGGTGCTTCTGGGATCGGATACCCGCCAAC 302  
 DB 1520 ---AlaProSerSerThrProGlyThrSerLeuAlaSerAlaSerProValPro----- 1536  
 QY 301 AGCACCCAGCTTCCGCTGAGCCCTCCACT-----CCTCGGC 263  
 DB 1537 -AlaProThrProValLeuAlaProSerSerThrGlnThrMetLeuProAlaProValPr 1556  
 QY 262 CTCCTCTCTCCAGTATGCGGTGACGACGACGACGACGACGACGACGACGACGACGAC 203  
 DB 1556 oSerProLeuProSerProAlaSerThrGlnThr-----LeuAlaLeuAlaPro-A 1573  
 QY 202 TGAGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 152  
 DB 1573 laLeuAlaProThr-LeuGlyGlySerSerProSerProSerGlnThrLeuSerLeuGlyThrGly 1592  
 QY 151 CAACCTCCGGAAGCTGCTGGTCTCTCTCACTCTCACTCTCACTCTCACTCTCACTCT 111  
 DB 1593 AsnProGlnGly--ProPheProThrGlnThrLeuSerLeuThrProAlaSerSerLeuV 1612  
 QY 110 -----ACACTTCACTCCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 68  
 DB 1612 alProThrProAlaGlnThrLeuSerLeuAlaProGlyProProLeuGlyProThrGlnT 1632  
 QY 67 GTGTCTCAGTAGCGAG-----AAAGAATCCAGACCTCAGGACCCGAGTCGC 20

Db 1632 hrLeuSerLeuAlaProAlaProLeuAlaProAlaSerProValGlyPro---AlaP 1651  
 QY 19 CGGCTCACAGC 9  
 DB 1651 roAlaHisThr 1654  
 RESULT 32  
 ID ABG21565 standard; Protein; 600 AA.  
 XX ABG21565;  
 XX 18-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #21556.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS85752.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 51924; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 600 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0.185 Length: 600  
 Score: 102.00 Matches: 50  
 Percent Similarity: 37.44% Conservative: 29



```
Db 662 hrThrTrpGlnAlaArgThrArgAlaAlaSerTrpSerProLeuGlyLeuAlaArgSerL 682
QY 344 CTGTGGGGTTCAGGTGCTTCTGGACTCGATACCCCAACAGCACCAGCTTCGCC 285
Db 682 euLeuProLeuAlaThrAlaSerProCysArgArgThrProThrGly----- 697
QY 284 TGAGCCCTCCACTCCTCGGCCCTCTCTCCCTAGTTATGGCCGTGACGAGACAAGAA 225
Db 698 -----SerGlyProThrTrpArgSerCysSerSerTrpGlyThrGluArgHisArgA 715
QY 224 GG-----TCCATGGCGGTGGACCGCATGAGAGGAGGACGACTTCGAGG 180
Db 715 snTrpArgGlnCysGlySerGlnGlySerAlaArgMetArgArg----- 729
QY 179 GCCTTCTCACTGAGTCACTGACACACACACCCCTCCGGAAGCTGCTGTTCTCTTCA 120
Db 730 -----ThrArgAlaAlaSerTrp-----T 736
QY 119 CTCCTCACTCACTTCAACTCCAGTTGGATCTGCCTGTGGACCTACCCG---CCGTGTC 63
Db 736 hrSerThrTrpGluAlaSerSerThrCysCysTrpTrpProTrpGlyTrpProCyst 756
QY 62 TCAGTAGCGGAGAGAAGATCCAGACCTCAGG-----ACCGAGTCCCGGCTCAC 12
Db 756 rpSerSerProGlySerThrTrpSerThrGlySerCysAlaThrArg-CysProThrHis 775
QY 11 AGCTCC 6
Db 776 ProSer 777

RESULT 34
AAY83167
ID AAY83167 standard; Protein; 87 AA.
XX AC AAY83167;
XX AC
XX AC
DT 24-JUL-2000 (first entry)
DE PAGE1 polypeptide.
XX KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
XX KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
XX KW cytotoxic T lymphocyte; immune response; antibody; drug delivery;
XX KW immunoconjugate.
XX OS Homo sapiens.
XX PN WO200012706-A1.
XX PD 09-MAR-2000.
XX PF 31-AUG-1999; 99WO-US20046.
XX PR 01-SEP-1998; 98US-0098993.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Pastan I, Brinkmann U, Vasmatazis G, Lee B;
XX DR WPI; 2000-237869/20.
XX PT
XX PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
XX PT lymphocyte response and for raising antibodies which can be used to
XX PT detect the presence of PAGE-4 in cell samples or body tissues
XX PS Disclosure; Figure 1b; 63pp; English.
XX CC PAGE-4 is a gene preferentially expressed in normal male and female
XX CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
XX CC and placenta, as well as in prostate cancer, testicular cancer and
XX CC uterine cancer. This expression pattern makes it a target for
XX CC diagnosis and for vaccine based therapy of such neoplasms.
```

```
CC An isolated PAGE-4 peptide which induces a cytotoxic T
CC lymphocyte response when bound to a major histocompatibility complex
CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC immunogenic compositions to raise a cytotoxic T lymphocyte response
CC against cells expressing PAGE-4 including cancer cells of the
CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC PAGE-4 peptide fragments can also be used in these compositions.
CC Antibodies against PAGE-4 and its peptide fragments can be used in
CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC samples or body tissues. The presence of PAGE-4 in tissues which are
CC not related to reproduction can be indicative of the spread of
CC cancerous reproductive tissue. PAGE-4 can also be used to raise
CC antibodies which are then used as the targeting group of
CC immunoconjugates comprising toxins used in therapeutic applications.
CC This has applications for drug delivery systems. This sequence is
CC the PAGE1 polypeptide which shares sequence similarity with
CC PAGE-4.
XX SQ Sequence 87 AA;
XX
Alignment Scores:
Pred. No.: 0.148 Length: 87
Score: 100.00 Matches: 28
Percent Similarity: 44.30% Conservative: 7
Best Local Similarity: 35.44% Mismatches: 25
Query Match: 7.25% Indels: 19
DB: 21 Gaps: 1
US-09-658-824-808 (1-781) x AAY83167 (1-87)
QY 441 GGAGAGCCCCAAAAGAACAGCAGCAGCTGAAAGTCGGGATCCTACACCTGGCAGCAG 500
Db 28 GlyGluSerGlnGluGluProProThrAspAsnGlnGlyPro----- 42
QY 501 ACAGAGAAGATCAGGATACAGCTGAGATCCCAGTCGCACATGGAAGGTGATCTGCAA 560
Db 43 -----AspMetGluAlaPheGlnGln 49
QY 561 GAGCTGCATCAGTCAAAACACCGGGATAAATCTGGATTTGGTTCGGCTCAAGGTGAA 620
Db 50 GluLeuAspLeuGlyThrArgSerGluArgGlyAsp-GlySerAspVallysGluLys 69
QY 621 GATAATACCTAAAGAGAACACTGTAAATGCCAGAGCAGGTGAAGAGCAACCA 675
Db 69 sThrProAsnProLysHisAlaLysThrLysGluAlaGlyAspGlyGlnPro 87
RESULT 35
AAM95890
ID AAM95890 standard; Protein; 203 AA.
XX AC AAM95890;
XX AC
XX AC
DT 21-NOV-2001 (first entry)
XX DE
XX DE Human reproductive system related antigen SEQ ID NO: 4548.
XX KW Human; reproductive system related antigen; reproductive system disorder;
XX KW cancer; gene therapy.
XX OS Homo sapiens.
XX PN WO200155320-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01339.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
```



CC including cancer. The present sequence is a protein of the invention.  
XX SQ Sequence 203 AA;

## Alignment Scores:

Pred. No.: 0.201 Length: 203  
Score: 100.00 Matches: 48  
Percent Similarity: 39.39% Conservative: 17  
Best Local Similarity: 29.09% Mismatches: 53  
Query Match: 7.25% Indels: 47  
DB: 22 Gaps: 8

US-09-658-824-808 (1-781) x AAW5890 (1-203)

QY 128 AACACGAGCTCCGAGGCTGTGTGTCAGTACAGTACAGAGGCGCTCGAAG 187  
DB 36 HSGHAlaGlyAspGlyClyArg-----LeuArgLeuLeuArgAlaAlaLys 51  
QY 188 TCGTCGCTCCCTCTCATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAA 247  
DB 52 -----GlyAlaAspPro----- 55  
QY 248 CTAGGAGGAGGCGGCGGAGGAGTGGAGGGCTCAGCGGAAGCTGGGTCTTGGG 307  
DB 56 ---GlyThrGlnArgIleArgGlyValHisGlyArgHisArg----- 68  
QY 308 GGTATCCGAGTCCACAGACCTCGAACCCCGACAGAGGATTCGTGGACTCCCGACAGCGG 367  
DB 69 -----ProGlyArgGlyThrGluSerGlu-----ProLysGly 79  
QY 368 GACACGAGGAGGCGGATGAGCGACACACACAAACACAGACACAGCCAGTCCCA 427  
DB 80 GluSerPheAspSerGlyValSerSer-IleGlyThrGluProAspSerValGluG1 99  
QY 428 GGAG-----CCGAGTAATGGAGCGCCCAAAAGAAAGAACACGAC 466  
DB 99 nGlnPheGlyProGlyAlaAlaArgAspSerGlnAlaGluProThrProGluGlnAl 119  
QY 467 AGCTGAAGTCGGGATCCTACACCTGGGCGACGACAGAGAAGATGATACAGTGA 526  
DB 119 aAlaGluAla---ProAlaGluGlyGlyProGlnThrAsnGlnLeuGluThrGlyAlaSe 138  
QY 527 GATCCCGAGTGGC-----GACATGAAGGTGATCTGCAAGAGCTGCATCAGTCAACAC 580  
DB 138 rSerProGluArgSerAsnGluValGluMetAspSerThrValIleThrValSerAsnSe 158  
QY 581 CGGGGATAAATCT 593  
DB 158 rSerAspLysSer 162

## RESULT 36

AAW57046  
ID AAW57046 standard; Protein; 411 AA.

XX  
XX  
AC AAW57046;

DT 20-AUG-1998 (first entry)

DE Mouse apoptosis inducing receptor.

XX KW Apoptosis inducing receptor; AIR protein; mouse; cell death regulator;  
KW Type I transmembrane protein; tumour cell death; autoimmune disease;  
KW therapy.

XX OS Mus sp.

XX PN W09814565-A1.

XX PD 09-APR-1998.

XX PF 03-OCT-1997; 97W0-US17876.

XX PR 04-OCT-1996; 96US-0044456.

XX

PA (IMMV ) IMMUNEX CORP.

XX PI Perkins PA;

XX WI PI; 1998-240077/21.

XX N-PSDB; AAV28701.

XX DNA encoding apoptosis inducing receptor - which is Type I  
XX transmembrane protein, useful for regulating cell death

PS Claim 16; Page 33-35; 45pp; English.

XX This sequence is the mouse apoptosis inducing receptor (AIR) of the  
XX invention. AIR is a Type I transmembrane protein, soluble forms of which  
XX can be used to regulate cell death in a therapeutic setting. Soluble AIR  
XX can also be used in vitro to block apoptosis or AIR-expressing cells, or  
XX to screen agonists or antagonists of AIR activity. The cytoplasmic domain  
XX of AIR can be used to develop assays for inhibitors of AIR-induced cell  
XX death, which is useful to regulate cell death in a therapeutic setting as  
XX well as in vitro. Agonists of AIR activity can be used to kill tumour  
XX cells that express AIR, or T cells expressing AIR in autoimmune diseases.

SQ Sequence 411 AA;

## Alignment Scores:

Pred. No.: 0.325 Length: 411  
Score: 99.00 Matches: 63  
Percent Similarity: 33.98% Conservative: 24  
Best Local Similarity: 24.61% Mismatches: 70  
Query Match: 6.94% Indels: 99  
DB: 19 Gaps: 14

US-09-658-824-808 (1-781) x AAW57046 (1-411)

QY 677 TGTGGTCTCTTCCACCTGCTTCTGGCATTTTACAGTGTCTCTTTAGGTATTATCTTC 618

DB 135 CysGlyLysSerSerProPheSer-----CysValProCysGly-AlaThrTh 150

QY 617 ACCTTGACGCCGGAACCCCAATCCA---GATTATCCCGGTGTT-----CA 573

DB 150 rProValHisGluAlaProThrProArgProCysLeuProGlyPheThrIleArgGlyAs 170

QY 572 CTGATGACAGCTTGTGCAGATCACCTTCCATGTCGCGCACTGGGATCTCAGCTGTATCTCTG 513

DB 170 nAspCysThrSerCys-----ProThrGlyPheSerSerVal----- 182

QY 512 ATCTTCTTCTCTGCTCCCGAGGTAGATCCGACTTTTACAGTGTCTGTTCTTCTTT 453

DB 183 -----CysProLysAlaCysThrAlaValCysGlyTrpLysGlnMetPh 197

QY 452 TTGGGGCTCTCCATTACTGGGCTCTCTGGGACTGGCTGTGTTGTTGTTGTTGTTGTTGTT 393

DB 197 eTrpValGlnValLeuLeuGlyVal---AlaPheLeuPheGlyAlaIleLeuIleCysAl 216

QY 392 CGCTCATGCGCTCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 333

DB 216 a-----TyrCysArgTrpG1 221

QY 332 CAGGTGC-----TTCCTGGGACTCGGATACCCCAACAGC---ACCCAGCTTCGCTCGA 282

DB 221 nProCysLysAlaValValThrAlaAspThrAlaGlyThrGluThrLeuAlaSerProG1 241

QY 281 GCCCTCCACTCTCCGCGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 222

DB 241 nThrAlaHisLeuSerAla----- 247

QY 221 CCATGGGCGTGGCACCAGCATGAGAGGACGACGACTTCGAGGGCCTTCTCACTCTGAGTC 162

DB 248 -----SerAspSe 250

QY 161 ACTGACCACACACCCCTCCGGAAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102

US-09-658-824-808 (1-781) x ABB60019 (1-773)

Qy 221 GACCTCTTGTCTGTCACGGCCCAACTAGGAGGAAGAGGGCCGAGAGTGAGGGG 280  
||||| :::|||  
Db 658 AspLeuLeuThrGlnAsnArgPheAlaLeuArgIysThrGlyArgSerLeuGluIle 677  
Qy 281 CTCAGCGCAAGCTGGGTGCTGTTGGGGTATCCGAGTCCCAGAAGCACCCTGGAACCCG 340  
||| ::::: ||| |||  
Db 678 ProGluGlnValAlaProLysGlyAspLeuGlnLysGluGluLysProLysGluGlu 697  
Qy 341 ACAGAAGATTCTGCATCCCCACAGCGGACAGGAGGACGGCATGACGCGACACAC 400  
::: |||||::: ::: ::: ::::  
Db 698 GluGlnLysGluLysLeuProLysGluGluValGln-----LeuGluGluIleLys 714  
Qy 401 AAACACAGAACCCACACAGC-----CAGTCCCAGGAGGCCAGTAATGGAGAGCCCA 451  
||| ||||| ||| :::||| |||  
Db 715 LysGlu-GluProGlnLysGluGluLeuGlnLysGluGluProGlnLysGluGluProG 734  
Qy 452 AAAAGAAGAACCCACGACTGAA-----AGTCGGGATCCTACACCTGGGCGACGACAGA 505  
||||| ||| ::::: ::::: ::::  
Db 734 nLysGluGluProArgLysGluGluProGlnLysGluGluProGlnLysGluGluProG 754  
Qy 506 AGAAGATCAGGATACAGCTGAGATCCGAGTCGCGACATGGAAGGTGATCTGCAAGAGCT 565  
: ::::: ||| |||  
Db 754 nLysGluGluProLysValGluThrPro-----GlnProLe 766  
Qy 566 GCATCAGTCA 575  
| |||||  
Db 766 uGluGlnSer 769

RESULT 38  
ABB67402  
ID ABB67402 standard; Protein; 821 AA.  
XX XX ABB67402;  
XX XX  
DT DT 26-MAR-2002 (first entry)  
DE DE Drosophila melanogaster polypeptide SEQ ID NO 28998.  
KW KW Drosophila; developmental biology; cell signalling; insecticide;  
KW KW pharmaceutical.  
OS OS Drosophila melanogaster.  
PN PN WO200171042-A2.  
XX XX 27-SEP-2001.  
PD PD  
PF PF 23-MAR-2001; 2001WO-US09231.  
PR PR 23-MAR-2000; 2000US-191637P.  
XX XX 11-JUL-2000; 2000US-0614150.  
XX XX (PEKE ) PE CORP NY.  
PA PA  
XX XX  
PI PI Venter JC, Adams M, Li PWD, Myers EW;  
DR DR WPI; 2001-656860/75.  
DR DR N-PSDB; ABL11505.  
XX XX  
PT PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -  
PT PT  
PS PS Disclosure; SEQ ID NO 28998; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA





QY 401 TGCTGTGT----- 393  
 Db 307 CysValCysValAspCysLeuCysSerSerLeuHisSerProArgArgGlyAlaGln 326  
 QY 392 -----CGCTCATCGCTCCCTCT----- 375  
 Db 327 TrpProArgThrCysGlnGluAsnLysAlaGlyAlaAlaArgTrpAsnAspGluThrAla 346  
 QY 374 -----CCTGGTCCCGTC--- 363  
 Db 347 HisLeuAlaAlaGluGlyArgAlaGluLysSerProGlnProProAlaProLeuSer 366  
 QY 362 -----TGGGAGTCCAGAAATCTTCTGTCGGGTCCAGGTGCTT 324  
 Db 367 CysLeuLeuValAspLeuProSerGlyProGluGluAlaGluAspHisPheLeuAlaArg 386  
 QY 323 CTGGGACTGGATACCCCAACAGCACCCAGCTTCGCTGAGCCCTCCACTCCPCGCGC 264  
 Db 387 ProGlyLeuGlyTyrCysArgArgArg-ProArgSerArgAlaProProAlaAlaAlaPr 406  
 QY 263 CCTCCTTCCTCCCTAGTTATGCGCGTGACGAGACAAAGGTCATGCGGTGGCACCGC 204  
 Db 406 oCysPheProPro-----SerArgLeuHisValGlnGlyLeuAl 419  
 QY 203 ATGAGAGGACGACGACTTCGAGGGCTTCTCACTCTGAGTCACTGACCACACAAGCC 146  
 Db 419 a-----AlaArgSerArgProLeuThrSerAlaPro 429  
 RESULT 40  
 AAE02397  
 ID AAE02397 standard; Protein; 522 AA.  
 XX  
 AC AAE02397;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Canine retinitis pigmentosa GTPase regulator (RPGR) protein.  
 XX  
 KW Dog; X-linked progressive retinal atrophy; XLPA; genetic marker;  
 KW retinitis pigmentosa GTPase regulator; RPGR; Siberian Husky; Samoyed;  
 KW Miniature Schnauzer.  
 XX  
 OS Canis familiaris.  
 XX  
 PN WO200138578-A1.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PF 21-NOV-2000; 2000WO-US31940.  
 XX  
 PR 24-NOV-1999; 99US-0167365.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 XX  
 PI Aguirre GD, Acland GM, Zhang Q, Ray K, Zeiss CJ;  
 XX  
 DR WPI; 2001-367707/38.  
 DR N-PSDB; AAD07130.  
 XX  
 PT Identifying dogs with or carrying X-linked progressive retinal atrophy  
 PT by detecting retinitis pigmentosa GTPase regulator gene mutation,  
 PT useful when breeding Husky, Samoyed and Miniature Schnauzer -  
 XX  
 PS Claim 72; Page 31-32; 88pp; English.  
 XX  
 CC The invention relates to a method for identifying dogs which are  
 CC genetically normal, are carriers of, or are affected with X-linked  
 CC progressive retinal atrophy (XLPA), by testing a biological sample with  
 CC genetic markers that co-segregate with a XLPA gene locus. The invention  
 CC also relates to canine retinitis pigmentosa GTPase regulator (RPGR), RPGR  
 CC mutants and their corresponding nucleic acid molecules. The mutated RPGR  
 CC genes are responsible for the XLPA in dogs. Methods are used to select  
 CC dogs for breeding so that dogs carrying the mutated locus are eliminated

CC from the breeding stock. The method particularly applies to Siberian  
 CC Husky, Samoyed and Miniature Schnauzer breeds, or any other breed where  
 CC the disease is X-linked. XLPA1 type is identified in Siberian Huskies,  
 CC and Samoyeds, while XLPA2 type is identified in Miniature Schnauzers.  
 CC The present sequence is Canine retinitis pigmentosa GTPase regulator  
 CC (RPGR) protein which is obtained from the normal open reading frame  
 CC (ORF) 15 of canine RPGR cDNA.  
 XX

SQ Sequence 522 AA;

Alignment Scores:

Pred. No.: 0.397 Length: 522  
 Score: 98.50 Matches: 52  
 Percent Similarity: 37.17% Conservative: 32  
 Best Local Similarity: 23.01% Mismatches: 112  
 Query Match: 7.14% Indels: 30  
 DB: 22 Gaps: 7

US-09-658-824-808 (1-781) x AAE02397 (1-522)

QY 69 GCGGTAGTCCACAGGAGATCCAACTGGGAGTTGAAG----- 107  
 Db 224 AlaGlyGlyGluGluGlyGluAlaGluTrpGluGlyLysGlyValGluGlyGlu 243  
 QY 108 TGTGAGTGAGAGTGAAGAGGAAACCAGCAGGCTTCGGAGGGTGTGTGTCAGTGACTCA 167  
 Db 244 GlyGluValGluGluGluGlyGluGluValGluGluGlyGlyGlyGlyGlyGly 263  
 QY 168 GAG---TGAGAGGGCCCTCGAAGTGTCTCCTCCTCATCGGTGCCAGCCCATGGACC 224  
 Db 264 GluValGluGluGlyGluAlaGluGluAspAlaGluGluGluGluArgGlu----- 281  
 QY 225 TTCTTGTCTCAGCGGCCATACTAGGAGGAAGAGG----- 263  
 Db 282 ---ValGluGluGluGlyGlyGlyGlyAsnLysArgGluTrpGlyGlyGluGlu 300  
 QY 264 -----GCCGAGGAGTGGAGGGCTCAGCGAGCTCAGCGAGCTGGGTGCTGTG 305  
 Db 301 GluGlyGluGluGluGluArgGluAsnGluArgGluGluGluGluGlyGlyGluAla 320  
 QY 306 GGGGTATCCGAGTCCCAAGACACCTCGAAGCCCGACAGAGATTCTGCACTCCCGACAC 365  
 Db 321 GlyGlyGlyGluGluGluGlyGluAlaGlyGlnGluGlyGlyGlyGlyGlyGly 340  
 QY 366 GGGACACGAGAGGAGCGCATGAGCGACACACAAACACACAGACACACAGCCAGCTCC 425  
 Db 341 GlyGluGluValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 360  
 QY 426 CAGGAGCCAGTAATGGAGAGCCCAAGAAAGAACAGCAGCTGAAAGTCGGATCCT 485  
 Db 361 GluGlu-----GlyGluGluGlyGlyGluGluGlyGlyGlyGlyGlyGlyGly 377  
 QY 486 ACACCTGGCAGCAGACAGAAGATCAGGATACAGCTGAGATCCCATGCCGACATG 545  
 Db 378 GlyArgGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 397  
 QY 546 GAAGGTGATCTCAA---GAGCTGCATCAGTCAACACCGGGGATAAATCTGGATTGGG 602  
 Db 398 GluGlyAspLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 417  
 QY 603 TTCCGGCTCAAGGTGAAGATAAATACCTAAAGAGGAACACTGTAAATGCCAGAGCAGG 662  
 Db 418 ValArgGluArgGluGluLysAspThr-GluGluGlyGly---LysTyrGluGluThrGlu 436  
 QY 663 TCAAGAGCAACCAAA 678  
 Db 436 yAspGluGluSerGlu 441

Search completed: October 18, 2002, 09:44:34  
 Job time : 66 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on:            October 18, 2002, 09:39:56 ; Search time 50 Seconds  
                               (without alignments)  
                               5404.362 Million cell up

Title: US-09-658-824-808

Perfect score: 1380  
 Sequence: 1 gcggcgagctatgagccg.....cttcacccaaaaaaaaa 781

Scoring table: BLOSUM62

xgapop	10.0	, xgapext	0.5
ygapop	10.0	, ygapext	0.5
fgapop	6.0	, fgapext	7.0
delop	6.0	, delext	7.0

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters:	1124444
--	---------

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%  
Listing first 45 summaries

**Command line parameters:**

```

-MODEL=framet_n2p_model_-dev_xlp
-O/cq12_1/USTPO.spool/US09568824/runat_l8102002_093947_9942/app_query_fasta_-1.967
-DB=SPRMBL19_OEFT-END-1 SUFFIX=rbstcct -MINMATCH=0 -L_OPCU=0 -L_OPEXT=0
-UNITS=bits_start-1 -fast-1 MATRIX=blosum62 TRANS=human40 cdi -LIST=45
-DLOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09568824_0CN1_1_88 #runat_l8102002_093947_9942 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NSG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=1 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YCAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database :

```

SPTEMB1_19:*
1:  sp.archaea:*
2:  sp.bacteria:*
3:  sp.fungi:*
4:  sp.human:*
5:  sp.invertebrate:*
6:  sp.mammal:*
7:  sp.mhc:*
8:  sp.organelle:*
9:  sp.phage:*
10: sp.plant:*
11: sp.potent:*
12: sp.virus:*
13: sp.invertebrate:*
14: sp.unclassified:*
15: sp.rvirus:*
16: sp.bacteriap:*
17: sp.archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	425	30.8	81	4 Q969J6	Q969J6 homo sapien

2	233	16.9	111	4	Q96GT9	Q96GT9	homo sapien
3	109	7.9	146	4	Q9BS57	Q9BS57	homo sapien
4	109	7.9	888	2	Q93GW8	Q93GW8	streptomyce
5	108	7.8	1003	12	Q91LX9	Q91LX9	kaposi's sa
6	106.5	7.7	976	12	Q9DUN0	Q9DUN0	kaposi's sa
7	106	7.7	398	4	Q96D06	Q96D06	homo sapien
8	106	7.7	1045	10	Q94HT6	Q94HT6	oryza sativ
9	105.5	7.6	528	4	Q9P0W9	Q9P0W9	homo sapien
10	104.5	7.6	355	2	Q918W2	Q918W2	mycoplasma
11	104.5	7.6	384	2	Q918V9	Q918V9	mycoplasma
12	104.5	7.6	538	11	Q9RT15	Q9RT15	mus musculus
13	104	7.5	319	16	Q9RV01	Q9RV01	deinococcus
14	103.5	7.5	498	11	Q9CSL7	Q9CSL7	mus musculus
15	103	7.5	246	11	Q63134	Q63134	rattus norv
16	103	7.2	413	11	Q99MM1	Q99MM1	mus musculus
17	103	7.5	643	5	Q9U229	Q9U229	caenorhabdi
18	102.5	7.4	1036	12	Q9UDW3	Q9UDW3	kaposi's sa
19	102.5	7.4	1162	12	Q9B148	Q9B148	kaposi's sa
20	102.5	7.2	2971	4	Q9Y5L9	Q9Y5L9	homo sapien
21	102	7.2	350	8	Q33572	Q33572	trypanosoma
22	102	7.2	1285	4	Q96JL1	Q96JL1	homo sapien
23	100.5	7.3	1054	10	Q9SN08	Q9SN08	oryza sativ
24	100	7.2	480	5	Q27033	Q27033	theileria p
25	99.5	7.2	181	10	Q94E85	Q94E85	oryza sativ
26	99.5	7.0	325	4	Q15627	Q15627	homo sapien
27	99	7.2	116	4	Q9UEU5	Q9UEU5	homo sapien
28	99	7.2	630	11	Q99JV6	Q99JV6	mus musculus
29	99	7.2	773	5	Q76871	Q76871	drosophila
30	99	7.2	773	5	Q94VX0	Q94VX0	drosophila
31	99	7.2	821	5	Q917X0	Q917X0	drosophila
32	98.5	7.1	405	10	Q41516	Q41516	trititum ae
33	98.5	7.1	405	10	Q9SRY0	Q9SRY0	trititum ae
34	98.5	7.1	847	6	Q9SKP7	Q9SKP7	canis famli
35	98.5	7.1	1012	10	Q949B8	Q949B8	oryza sativ
36	98.5	7.1	1422	6	Q95K04	Q95K04	canis famli
37	98	7.1	299	4	Q13059	Q13059	homo sapien
38	98	7.1	1089	12	Q40947	Q40947	kaposi's sa
39	98	7.1	1245	4	Q43636	Q43636	homo sapien
40	98	7.1	1251	4	Q9UMG2	Q9UMG2	homo sapien
41	98	6.9	1668	4	Q15026	Q15026	human herpe
42	97.5	7.1	296	12	Q69118	Q69118	human herpe
43	97	7.0	345	5	Q9VJ23	Q9VJ23	drosophila
44	97	6.8	417	5	Q9V6Q3	Q9V6Q3	drosophila
45	97	6.0	457	11	Q88344	Q88344	rattus norv

## ALIGNMENTS

RESULT 1	
Q969J6	
ID	PRELIMINARY; PRT; 81 AA.
AC	
DT	Q969J6;
CD	01-DEC-2001 (TREMBLrel. 19, Created)
DD	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	CANCER/TESTIS ASSOCIATED PROTEIN XAGE-1B (HYPOTHETICAL 9.1 KDA DE PROTEIN) (9 KDA CANCER/TESTIS ASSOCIATED PROTEIN).

OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

RP SEQUENCE FROM N.A.  
RC TISSUE=MELANOMA METASTASIS;  
RA Zendenan A.J.W., van Kraats A.A., Weidle U.H., Rulter D.J.  
RZ Van Muijen G.N.P.;  
RT "XAG1-1b, a novel cancer/testis-associated gene, induced  
RT metastases";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG CARCINOMA, AND LARGE CELL UNDIFFERENTIATED;  
RA  
RZ  
RT  
RL

RA Strausberg R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-MELANOMA METASTASIS;  
 RA Zenden A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.J.,  
 Van Muijen G.N.P.;  
 RT "Ctp9, a novel cancer/testis-associated gene, induced in melanoma  
 metastases";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ400997; CAC38108.1; -;  
 DR EMBL; BC009538; AA09538.1; -;  
 DR EMBL; AJ290447; CAC38107.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 81 AA; 9078 MW; C73337C4C94C01D1 CRC64;

Alignment Scores:  
 Pred. No.: 1.16e-32 Length: 81  
 Score: 425.00 Matches: 81  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 30.80% Indels: 0  
 DB: 4 Gaps: 0

US-09-658-824-808 (1-781) x Q969J6 (1-81)

QY 439 ATGGAGAGCCCAAAAGAGAACAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGC 498  
 Db 1 MetGluSerProLysLysLysAsnGlnLeuLysValGlyLeuHisLeuGlySer 20  
 QY 499 AGACAGAGAGATCAGGATACAGCTGAGTCCAGTCGGGACATGGAAGGTATGTC 558  
 Db 21 ArgGlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysValIleCys 40  
 QY 559 AAGAGCTCATCAGTCACAAACACCGGGGATAATCTGGAATTTGGTTCGGCGTCAAGGTG 618  
 Db 41 LysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuGlySerGlyValLysVal 60  
 QY 619 AAGATATACCTAAAGAGAACACCTGTAAATGCCAGAGCAGGTGAAGACCAACCAA 678  
 Db 61 LysIleIleProLysGluGluHisCysLysMetProGluAlaGlyGluGluGlnProGln 80  
 QY 679 GTT 681  
 Db 81 Val 81

RESULT 2

Q96GT9 ID Q96GT9 PRELIMINARY; PRT; 111 AA.  
 AC Q96GT9;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE SIMILAR TO G ANTIGEN 8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PLACENTA, AND CHORIOCARCINOMA;  
 RA Strausberg R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC009232; AA09232.1; -;  
 SQ SEQUENCE 111 AA; 12354 MW; 38C9332C5BA0E14 CRC64;

Alignment Scores:  
 Pred. No.: 4.54e-14 Length: 111  
 Score: 233.00 Matches: 48  
 Percent Similarity: 74.12% Conservative: 15  
 Best Local Similarity: 56.47% Mismatches: 21  
 Query Match: 16.88% Indels: 1  
 DB: 4 Gaps: 0

US-09-658-824-808 (1-781) x Q96GT9 (1-111)  
 QY 429 GAGCCAGTAATGGAGAGCCCAAAAGAGAACAGCAGCTGAAAGTCGGATCTTACA 488  
 Db 28 GluProThrAspGluGluProLysGluGluLysProThrLysSerArgAsnProThr 47  
 QY 489 CTGGGACGACAGACAGAGATCAGGATACAGTCCAGTCCGCGACATCGAA 548  
 Db 48 ProAspGlnLysArgGluAspGlnGlyAlaAlaGluIleGlnValProAspLeuGlu 67  
 QY 549 GGTGATCTGCAAGAGCTCATCAGTCAAAACACCGGGGATAAATCTGGATTTGGTTCGG 608  
 Db 68 AlaAspLeuGlnLeuLysGlnThrLysThrGlyAsp-GlyCysGluGlyGlyThrAs 87  
 QY 609 CGTCAAGGTGAAGATAATACCTAAAGAGAACACCTGTAAATGCCAGAGCAGGTGAAGA 668  
 Db 87 pValLysGlyLysIleLeuProLysAlaGluHisPheLysMetProGluAlaGlyGluGlu 107  
 QY 669 GCAACCAACAGTT 681  
 Db 107 YLysSerGlnVal 111

RESULT 3

Q9BSS7 ID Q9BSS7 PRELIMINARY; PRT; 146 AA.  
 AC Q9BSS7;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DE G ANTIGEN, FAMILY B, 1 (PROSTATE ASSOCIATED).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-MELANOMA;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC004861; AA04861.1; -;  
 SQ SEQUENCE 146 AA; 16150 MW; B6C7BA94D898DCB2 CRC64;

Alignment Scores:  
 Pred. No.: 0.0468 Length: 146  
 Score: 109.00 Matches: 33  
 Percent Similarity: 50.00% Conservative: 15  
 Best Local Similarity: 34.38% Mismatches: 41  
 Query Match: 7.90% Indels: 7  
 DB: 4 Gaps: 2

US-09-658-824-808 (1-781) x Q9BSS7 (1-146)  
 QY 423 TCCAGAGCCAGTAATGGAGAGCCCAAAAGAGAACAGCAGCTGAAAGTCGGAT 482  
 Db 20 SerGluGluSerSerAspGluGlnProAspGluValGluSerProThrGlnSerGlnAsp 39  
 QY 483 CCTACACCTGGGCAGCAGACAGAGATCAGGATACAGTCCAGTCCCGCGAC 542  
 Db 40 SerThrProAlaGluGlu---ArgGluAspGluGlyAlaSerAlaAlaGlnGlyGlnGlu 58  
 QY 543 ATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATAAATCTGGATTTGGG 602  
 Db 59 ProGluAlaAspSerGlnGluLeuValGlnProLysThrGlyCysGluLeuGlyAsp-Gl 78  
 QY 603 TTCGGCGCTCAAGTGAAGATATACTAAAGAGGAAACACTGTAAATGCCAGAGCAGG 662  
 Db 78 YProAspThrLysArgValCysLeuArgAsnGluGluGlnMetLysLeuProAlaGluGlu 98  
 QY 663 TGAAGAG-----CAACCAACAGTTTAAATCAACACA 693  
 Db 98 YProGluProGluAlaAspSerGlnGlnValHisProLysThr 113

```
RESULT 4
Q93GW8      PRELIMINARY:      PRT;      888 AA.
ID   Q93GW8;
AC   Q93GW8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE   PUTATIVE LARGE SECRETED PROTEIN.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   *Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.*;
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070957; BAB69423.1; -.
SQ   SEQUENCE 888 AA; 93501 MW; 061EC2522217C15F CRC64;

Alignment Scores:
Pred. No.:      0.0618      Length:      888
Score:          109.00      Matches:      45
Percent Similarity: 35.95%      Conservative: 10
Best Local Similarity: 29.41%      Mismatches: 64
Query Match:      7.90%      Indels:      35
DB:              2      Gaps:      7

US-09-658-824-808 (1-781) x Q93GW8 (1-888)
QY 238 ACGCCATACTAGGAGGAGGCGGAGGAGTGGAGGCTCAGGCGAAGCTGGG 297
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 684 ThrArgLeuProAspHisSerGlyPheProGluTyrGlyGlyAla---GlyLeuLeuPro 702
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 298 TGCTGTGGGGTATCCGAGT-----CCGAGAACACCTGGAACCCCGACAGA 345
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 703 AlaAlaSerAspHisProAlaProAlaProGlyProAsnAlaSerTAlaProArg 722
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 346 AGATTCTGGACTCCCGAG-----ACGGGACACGAG----- 375
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 723 GluHisProAlaProGlnGlyHisArgThrGluProGlyGlyLeuGlyProGlnAlaGly 742
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 376 -----GAGGACCGCATGCGCACACACACAAAC-----ACAGAA 410
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 743 GlnProAlaHisProGlyAlaHisAlaGlyHisHisGlySerArgAlaGlyPheThrGlu 762
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 411 CCACACAGCCAGTCCAGGAGCCCGAGTAAATGGAGAGCCCGCCAAAGAGAACAGCAGCT 470
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 763 ProGlnSerGlyPheThrGluProArgSerGlyProProSerArgHisGluProGlyAla 782
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 471 GAAAGTCGGGATCCTACACCTGGGCGAGCACAGAGAAGATCAGGATACAGCTCAGATC 530
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 783 GlySerGlyGlyProAlaProAlaArgGlu-----GluAla 794
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 531 CCAGTGCAGCATGGAAAGGTGATGTG-----CAAGAGCTGCATCATCAACACC 581
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 795 ProAlaArgSerAlaAlaArgProLeuThrThrProGlnValLeuProGlnArgThrArg 814
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 582 GGGGATAAATCTGGATTTGGGTTCCGGCTCAAGGTGAA 620
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 815 GlyAlaSerLeuAlaGlnGlnLeuArgArgGluAlaGlu 827
Db 111 : : : : : 111 111 111 111 111 111 111 111 111

RESULT 5
Q91LX9      PRELIMINARY:      PRT;      1003 AA.
ID   Q91LX9;
AC   Q91LX9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OS   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99445611; PubMed=10515805;
RA   Gao S.J., Zhang Y.J., Deng J.H., Rabkin C.S., Flore O., Jenson H.B.;
RT   *Molecular polymorphism of Kaposi's sarcoma-associated herpesvirus
RT   (Human herpesvirus 8) latent nuclear antigen: evidence for a large
RT   repertoire of viral genotypes and dual infection with different viral
RT   genotypes.*;
RL   J. Infect. Dis. 180:1466-1476(1999).
DR   EMBL; AF360120; AAK50002.1; -.
SQ   SEQUENCE 1003 AA; 115517 MW; C20C43308B01A0A3 CRC64;

Alignment Scores:
Pred. No.:      0.0786      Length:      1003
Score:          108.00      Matches:      28
Percent Similarity: 40.57%      Conservative: 15
Best Local Similarity: 26.42%      Mismatches: 41
Query Match:      7.83%      Indels:      22
DB:              12      Gaps:      2

US-09-658-824-808 (1-781) x Q91LX9 (1-1003)
QY 261 AGGCGGAGGAGTGGAGGCGCTCAGGCGAAGCTGGGTGCTGTGGGGGTATCCGAGTCC 320
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 409 LysLysGluAspGluGluAspGlyGlyAspGlyAsnLysThrLeuSerIleGlnSerSer 428
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 321 CAGAAGCACCTGGAAACCCCGACAGAGAGATCTCGACTCCCGACGCGGACGAGAGGG 380
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 429 GlnGlnGlnGlnGluProGlnGlnGln----- 437
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 381 ACGGCATGAGCGACACACACAAACACAGAACACACAGCCAGTCCCGAGGAGCCAGTAAT 440
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 438 -----GluPro-----GlnGlnGlnGluProGlnGln 446
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 441 GGAGAGCCCCAAAAGAGAACACGAGCTGAAAGTCGGGATCCTACACCTGGGCGACGAG 500
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 447 GlnGluProGlnGlnGlnGluProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 466
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 501 ACAGAAGACATCAGGATACAGCTGAGATCCCGATGCGCGACGATGGAAGGTGATCTGCAA 560
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 467 GluProGlnGlnGlnGluProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 486
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 561 GAGCTGCATCAGTCAAAAC 578
Db 111 : : : : : 111 111 111 111 111 111 111 111 111
QY 487 GluProGlnGlnGlnAsp 492
Db 111 : : : : : 111 111 111 111 111 111 111 111 111

RESULT 6
Q9DUNO      PRELIMINARY:      PRT;      976 AA.
ID   Q9DUNO;
AC   Q9DUNO;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OS   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99445611; PubMed=10515805;
RA   Gao S.J., Zhang Y.J., Deng J.H., Rabkin C.S., Flore O., Jenson H.B.;
RT   *Molecular polymorphism of Kaposi's sarcoma-associated herpesvirus
RT   (Human herpesvirus 8) latent nuclear antigen: evidence for a large
RT   repertoire of viral genotypes and dual infection with different viral
RT   genotypes.*;
RL   J. Infect. Dis. 180:1466-1476(1999).
DR   EMBL; AF360120; AAK50002.1; -.
SQ   SEQUENCE 1003 AA; 115517 MW; C20C43308B01A0A3 CRC64;
```

[illegible]

Score:	106.00	Matches:	52
Percent Similarity:	37.74%	Conservative:	28
Best Local Similarity:	24.53%	Mismatches:	76
Query Match:	7.68%	Indels:	56
DB:	4	Gaps:	8
US-09-658-824-808 (1-781) x U96D06 (1-398)			
Qy	52	TCCGCTACTGAGACACGGCGGTAGGTCACACGACAGATCCAACCTGGAGTTCGAAGTGTG	111
Db	101	SerValThrGluLeu---GlnThrHisProGluLeuAspThrAspGlyAspGlyAlaLeu	119
Qy	112	AGTGAGAGTGAA-----GAGGAACACGACGAGCTTC	141
Db	120	SerGluAlaGluAlaGlnAlaLeuSerGlyAspThrGlnThrAspAlaThrSerPhe	139
Qy	142	CGGAGGGTTGTGGTCACTGAGTGTGAGAGGCCCTCGAAGTGTGTCGTCCTCTCTC	201
Db	140	TyrAspArgValTpaAlaAla-IleArgAspLysTyrArgSerGlu-----	154
Qy	202	ATGCGGTGCCACGCCCATGGACCTTCTGTCTCGTCACGCCCACTAAGGGAGG--AA	258
Db	155	-----AlaLeuProThrAspLeuProThrProSerAlaProAspLeuThrGluProly	172
Qy	259	GGAGGGCCGAGAGTGGAGGGGCTCAGCGCAAGCTGGGTGCTGTTGGGGGTATCCGAGT	318
Db	172	sGluGluGln-----	175
Qy	319	CCGAGAAGCACTGGAAACCCGACAGAGAAGATTCTTGACTCCCGACAGCGGACGAGAG	378
Db	176	-ProProValProSerProThrGluGluGlu-----GluGluGluGluGluGlu	192
Qy	379	GGACGGCATGAGGCACACACACAAACACACACACAGCCAGTCCCAGGAGCCCACTA	438
Db	192	uGluGluAlaGluGluGluGluGluGluGluAspSerGluGluAlaProProProL	212
Qy	439	ATGGAGAGCCCAAAAAGAAAGAACACGACAGTGGGATCGGTACCTACAGTGGGACG	498
Db	212	euserProGlnProAlaSerProAlaGluGluAspLysMetProProTyrAspGluG	232
Qy	499	AGACAGAAGAATCAGGATACA-----GCTGAGA	528
Db	232	lnThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGlu	252
Qy	529	TCCGAGTCCGCGATGAAGTGCATCGCAAGCTGTCATCAGTCAACACCGGGGATA	588
Db	252	rgSerLeuLysAspMetGluGluSerIleArgAsnLeuGluGln-----G	267
Qy	589	AATCTGATTTGGTTCCGGCGTCAAGGTGAA	620
Db	267	luIleSerPheAspPheGlyProAsnGlyGlu	277
RESULT 8			
Q94HT6 PRELIMINARY; PRT; 1045 AA.			
ID	Q94HT6		
AC	Q94HT6		
DT	01-DEC-2001 (T-EMBLrel. 19, Created)		
DT	01-DEC-2001 (T-EMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (T-EMBLrel. 19, Last annotation update)		
DE	HYPOTHETICAL PROTEIN.		
OS	Oryza sativa (Rice).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Ehrhartoideae; Oryzeae; Oryza.		
OX	NCBI_TaxID=4530;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NIPPONBARE.		
RA	Du H., Minx P., Abbott A., Doebber A., de la Bastide M., Spiegel L.,		
RA	Nascimento L., Preston R., Kirchoff K., King L., Vil M.D., Baker J.,		
RA	Zutavern T., Santos L., Bell M., Miller B., Kuit K., Rodriguez S.,		
RA	Cunha D.M., Balija V., Shah R., Bahret A., O'Shaughnessy A.,		
RA	Palmier L., Yang C., Dedhia N., McCombe W.R.;		

\*Genomic Sequence for *Oryza sativa*, Nipponbare strain, clone  
 WT OSJNB0036B06, from Chromosome 10, complete sequence.\*;  
 RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 LR SRR000000.100000000.100000000.100000000.100000000.  
 DR EMBL: AC051624; AAK92563.1; -  
 SO SQUENCE 1045 AA: B8344297B81FEFF CRC64;

```

RP      SEQUENCE FROM N.A.
RA      Marcil A., Sevigny G., Thomas D.Y.;
RT      "Expression of Enzymatically Active Isoforms of Human Glucosidase 11
RL      in Insect Cells.";
RL      Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF144075; AAF66686.1; -.
DR      InterPro: IPR002048; EF-hand.
DR      InterPro: IPR000886; ER.target.
DR      InterPro: IPR002172; LDL_recept_.
DR      SMART: SM00192; LDLa; 1.
DR      PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR      PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
SQ      SEQUENCE 528 AA; 59425 MW; 8DAD9776037E78E CRC64;

Alignment Scores:
Pred. No.: 0.124 Length: 528
Score: 105.50 Matches: 52
Percent Similarity: 23.74% Conservatve: 28
Best Local Similarity: 24.53% Mismatches: 77
Query Match: 7.64% Indels: 55
DB: 4 Gaps: 8

```

DT 01-OCT-2000 (TREMURel. 15, last sequence update)  
DT 01-DEC-2001 (TREMURel. 19, Last annotation update)  
DE 87-KDA SURFACE LIPOPROTEIN (FRAGMENT).  
GN VLPAB7.  
OS Mycoplasma hyorhinis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2100;

**RESULT**

Db    Qy    Db	<div>637 GAACACTGTAAAATGCCAAGCAGGCTGAAGAACCAACCACAAGTTAAATGAAGACAAG             :: ::: :::: :: :: ::  283 lyThrThrAsnThrGlySerGlySerAsnSerGluSerGlyMetAsnSerGluLys</div>	<div>PRT: 384 AA  PRELIMINARY: PDB ID: P95088 Length: 13 Created: Accession: 13 Left sequence update) Accession: 13 Left annotation update)</div> <div>Q9LSV9 VLPAUT. OS Mycoplasma hyorhinis. OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes; CC Mycoplasmataceae; Mycoplasma. OX NCBI_TaxID=2100; KE [ ] ME [ ] RP SEQUENCE FROM N.A. RC STRAIN-SK76; RX MEDLINE-20138163; PubMed-10671459; RA Citti C., Watson-McKown R., Droesse M., Wise K.S.; RT "Gene Families Encoding Phase- and Size-Variant Surface Lipoproteins of Mycoplasma hyorhinis"; EL J. Bacteriol. 182:1356-1363(2000). DR EMBL: AF193880; AAF36549.1; - KW SIGNAL; Lipoprotein.</div>
----------------	--	---

FT FT SQ	<div>SIGNAL 1 29 POTENTIAL. CHAIN 30 384 87-KDA SURFACE LIPOPROTEIN. SEQUENCE 384 AA; 382226 MW; 04AIFOA98E3B79C CRC64;</div>
----------	---

82	SerGlySerAsnSerGluSerGlyMetAsnSerGluLysThrGluAsnThrGlnGlnSer	101
QY	76 GGTCCACAGGCAGATCCA	ACTGGAGTTGAAGTGTGAGTCAGAGTGAAAGAGAA-----129
Db	102 GluAlaProGlyThrAsnThrGlyAsnLysThrSerGluSerAsnSerGluSerSer	121
QY	130 -----CACAGAGCTCCCGAGGGTTGCTGCTCAGTCACTCAGAGTGAGAAGGCC	180
Db	122 ThrGlySerGlnAlaGlyThrThrAsnThrGlySerGlySerAsnSerGluSerGly	141
QY	181 CTC-----GAAGTCGTCGTCCTCATGCGGTGCCAGCCCATGGACCTTCTTGTCTC	234
Db	142 MetAsnSerGluLysThrGluAsnThrGlnGlnSerGluAlaProGlyThrLys	159
QY	235 GTCACGCCCAATACTAGGAGGAGGAGGCCCGCAGGA-----GTGAGGGGCTCAGGCG	288
Db	160 ---ThrGluAsnThrGlnGlnSerGluAlaProGlyThrLysThrGluAsnThrGlnGln	178
QY	289 AAGCTGGGGTGCTGTTGGGGGTATCCAGTCCACAGAACGACCCTGGAAACCCGACACAAGA	348
Db	178 nSer-----GluAlaProGlyThrAsnThrGlyAs	188
QY	349 T-----TCTGGACTCCCACAGCGGNACAGGAGAGGGAGCGGCA	TGAGGCACACACA399
Db	188 nLysThrThrSer-GluSerAsnSerGlySerSerThrGlySerGlnAlaGlyThrThr	208
QY	400 CAACACAGAACCCACACAGCAGCTCCCGAGGCCCCAGTAGTAATGAGAGCGCCCAAAAAGAG	459
Db	208 hrAsnThrGlySerGlySerAsnSerGluSerGlyMetAsnSerGluLysThrGluAsnT	228
QY	460 AACCCAGCAGCTGAAGTCGGGATCC	TACACCTGGGAGCAGACAGAAGAATCAGGATCA519
Db	228 hrGlnGlnSerGluAla-----ProGlyThrLysThrGluAsnThrGlnGln-	243
QY	520 CAGCTGAGATCCCGAGTCGCGGACATGGAAGGTGATCTGCAA---GAGTCGCATCACTCAA	576
Db	244 --SerGluAlaProGlyThrLysThrGluAsnThrGlnGlnSerGluAlaLeuGlyThrA	263
QY	577 ACACGGGGGATAAATCTGGATTGGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAG	636

Alignment Scores:				
Pred. No.:	0.148	Length:	384	
Score:	104.50	Matches:	63	
Percent Similarity:	39.83%	Conservative:	33	
Best Local Similarity:	26.14%	Mismatches:	112	
Query Match:	7.57%	Indels:	33	
DB:	2	Gaps:	9	
US-09-658-824-808 (1-781) x Q9L8V9 (1-384)				
Qy	16	GC CGG CAC TCG GTC CCG TCG AGT TCG GAT TCT TCT CCG C TACT CAG ACAG CGG CGG TA	75	
Db	111	Ser Gly Ser Asn Ser Glu Ser Gly Met Asn Ser Glu Lys Thr Glu Asn Thr Glu Asn Thr Glu Asn Thr	130	
Qy	76	GG TCC CAG CCG CAG ATC CA ACT CCG GAG TTG AAG TGT CAG TGT CAG AGT GA A GAG AA	129	
Db	131	Glu Ala Pro Gly Thr Asn Thr Gly Asn Lys Thr Thr Ser Glu Ser Asn Ser Glu Ser Ser	150	
Qy	130	----- -C CAG CAG C TCC CCG GAG GTT G TGT CAG T CAG T CAG T CAG AGT CAG AAG GCC	180	
Db	151	Thr Gly Ser Glu Ala Gly Thr Thr Thr Asn Thr Gly Ser Gly Ser Asn Ser Glu Ser Gly	170	
Qy	181	CTC ----- -G AAG TCG TCG TCC TCT C ATC GCG TGC CAG CCG CCA TCT TCT TGT CTC	234	
Db	171	Met Asn Ser Glu Lys Thr Glu Asn Thr Glu Asn Thr Glu Ala Pro Gly Thr Lys	188	
Qy	235	GT CAC GCG CCA TACT AGG AGG AAG CAG GCG CCG CAG A----- -GT C GAG GCG C T CAG GCG	288	
Db	189	--- Thr Glu Asn Thr Glu Asn Thr Glu Ala Pro Gly Thr Lys Thr Glu Asn Thr Glu Asn Thr	207	
Qy	289	AAG TGG GGT GCT GTT GGG GGT AT CCG AGT CCG CAG AAG CAG CACT GGA CCCC CAG CAG AAG A	348	
Db	207	nSer----- -Glu Ala Pro Gly Thr Asn Thr Lys Thr	217	
Qy	349	T----- -TCT GGA CTCCC AG CCG GAC CAG GAG GAG GCG CCA TGT CAG CAG CAC A CACA	399	
Db	217	nLys Thr Thr Ser -Glu Ser Asn Ser Gly Ser Thr Thr Gly Ser Glu Ala Gly Thr Thr Thr	237	
Qy	400	CA A C A C A G A A C C A C A C C A G T C C C C A G G C C C A G T A A T G A G A G C C C A A A A G A G	459	



```

Db 237 hrAsnThrGlySerGlySerAsnSerGluSerGlyMetAsnSerGluLysThrGluAsnT 257
QY 460 AACACAGACGCTGAAGTCTGGGATCTTACACCTGGGAGCAGACAGAGATCAGGATA 519
Db 257 hrGlnGlnSerGluAla-----ProGlyThrLysThrGluAsnThrGlnGln- 272
QY 520 CACTGAGATCCAGTGGCGGACATGGAAGGTGATCTGCAA---GAGCTCATCACTCAA 576
Db 273 --SerGluAlaProGlyThrLysThrGluAsnThrGlnGlnSerGluAlaLeuGlyThra 292
QY 577 ACACCGGGATAATCTGGATTGGTTCGGCGCTCAAGGTGAAGATAATCAATAAGAG 636
Db 292 snThrGlyAsnLysThrThrSerGluSerAsnSerGlySerSerThrGlySerGluAlaG 312
QY 637 GACACTGTAAATGCCAGAGCAGGTGAAGACCAACCAAGTAAATTAAGACAAAG 695
Db 312 lyThrThrAsnThrGlySerGlySerAsnSerGluSerGlyMetAsnSerGluLys 331
RESULT 12
ID Q9ET15 PRELIMINARY; PRT; 538 AA.
AC Q9ET15;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE RETINITIS PIGMENTOSA GTPASE REGULATOR (FRAGMENT).
GN RPR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Vervoort R., Lennon A., Bird A.C., Tulloch B., Axton R., Miano M.G.,
RA Meindl A., Meitinger T., Ciccodicola A., Wright A.F.,
RT "Mutational hot spot within a new RPR exon in x-linked retinitis
RT pigmentosa.";
RL Nat. Genet. 0:0-0(2000).
DR EMBL; AF286473; AAG00552.1; -.
FT NON_TER 1
SQ SEQUENCE 538 AA; 60333 MW; 549F3E092943FDE0 CRC64;

Alignment Scores:
Pred. No.: 0.156 Length: 538
Score: 104.50 Matches: 59
Percent Similarity: 38.60% Conservative: 29
Best Local Similarity: 25.88% Mismatches: 90
Query Match: 7.57% Indels: 51
DB: 11 Gaps: 10

US-09-658-824-808 (1-781) x Q9ET15 (1-538)
QY 109 GTGAGTGAGAGTGAAGAGGAACAGCAGGCTCCGGAGG-----GTTGTGTGTCAGTG 162
Db 1 IleProGluGlnGluGlyProGluAspSerGluGlyAsnValValGluGlnVal 20
QY 163 ACTCAGAGTCAGAGGCC---CTCGAAGTCGTCTGCTCCTCATCGGTCGCCAGGCCAT 219
Db 21 ValGlnAlaGlnLysGluAsnLeuGluPheGluGlyAspArgLysGluAlaLysAlaGlu 40
QY 220 GGACCTCTTGTCTGCTCAGC-----GCCATACTAGGAGGAAGAGGG 264
Db 41 AlaProSerAspValIleThrGluLysGluValSerGluArgGluSerGlyGly 60
QY 265 CCAGAGAGTGG-AGGGCTCAGCGAAGCTGGGGTGGTGTG----- 305
Db 61 GluArgGluAspArgSerGluGlyAspGlyAspGlnIleCysGluLysValSerLeuGlu 80
QY 306 -----GGGTATCCGAGTCCAGAGCAGCAGGACCTTGAACCCGA 341
Db 81 ThrGluHisLeuGlnArgAlaGlnGlyLysGlnGluArgLysLysGlyLysAspLysArg 100
QY 342 CAGAAG---ATTCTGGACTCCCCAGAC-----GGGACCCAGAGAGGAGC 383

```

```

Db 101 AlaArgCysIleLeuAspMetLysGluArgGluGluAspLysGlyTrpGluLysGlySer 120
QY 384 GCATGAGCGACACACAAACACAGAACACACAGCAGCTCC----- 425
Db 121 GluGlyGlyAspLysMetLysArgAspGluGlyAsnGlnGluLysArgLysLysGluMet 140
QY 426 CAGAGCCCGAGTAAATGAGAGCCCCAAAAGAAAGAACACAGCTGAAAGTCGGGATCCT 485
Db 141 GluLysArgAspAlaGlyAspGluArgSerGluGluGluGluGluGluPro 160
QY 486 ACACCTGGGCGCAGCAGACA-----GAAGAAGATCAGGATACAGCTGAGATCCCCAGTG 536
Db 161 GluGluGlyGlyLysGluGlyGlyGluGluGluGluGlyThrSerGluAspGlnSer 180
QY 537 CGGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATAAATCTGGA 596
Db 181 ArgLysAspGluGlyAspArgGlnGlu-----LysGluGlyArgArgGluGly 196
QY 597 TTTGGGTTCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAATGCCAGA 656
Db 197 LysGlyArgGlnGluAspGlyArgGluGlyTrp-LysGlu----- 209
QY 657 AGCAGGTGAAGAGCAACACCAA 678
Db 210 ---GlyGluGluGlnGluGln 215
RESULT 13
Q9RV01
ID Q9RV01 PRELIMINARY; PRT; 319 AA.
AC Q9RV01;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 33.1 KDA PROTEIN.
GN DR1229.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001971; AAF10810.1; -.
DR TIGR; DR1229; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 319 AA; 33082 MW; 7F0229C5DAAA2B61 CRC64;

Alignment Scores:
Pred. No.: 0.161 Length: 319
Score: 104.00 Matches: 49
Percent Similarity: 39.62% Conservative: 14
Best Local Similarity: 30.82% Mismatches: 57
Query Match: 7.54% Indels: 39
DB: 16 Gaps: 9

US-09-658-824-808 (1-781) x Q9RV01 (1-319)
QY 19 GCGACTCGGGTCCCTCGAGGTCTGGATTCTTCTCCGCTACTGAGACACGCGGGTAGGT 78
Db 159 GlySerSerGlyProSerSerGlyCysSerAlaAsnThrThrArgAlaArgHisAlaPro 178
QY 79 CCACGCGCAGATCCCACTGGGAGTTGAAGTGTGAGTGAGAGTGAAGAGGAACACGAGGC 138

```



Query Match:	7.46%	Indels:	33
DB:	11	Gaps:	6
US-09-658-824-808 (1-781) x Q63134 (1-246)			
QY 306	GGGTATCCGAGTCCAGACACGCTGGAAACCCCGACAGAAAGATTCTGGACTCCCGACAG	365	
DB 74	GlySerGluGluGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGluProPro	93	
QY 366	GGGACGAGGAGGAGCGGATGACGACACACACAAACACAGAACACACAGCCAGTCC	425	
DB 94	AlaThr-----SerGlySerGluGluGlnGlnGln	104	
QY 426	CAGGAGCCAGT-----AATGAGAGCCG	449	
DB 105	GlnGluProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySerGluGluGlu	124	
QY 450	-----CAAAAGAGAACACGCA---GCTGAAAGTCGGGATCTACACCTGGGCAGCAGACA	503	
DB 125	GlnGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySer	144	
QY 504	GAAGAGATCAGATACAGCTGAGATCCAGTCGCGACATGGAGGTGATCTGCAAGAG	563	
DB 145	GluGluGlnGlnGlnGlnGlnGlu-----SerThrGlnAlaGluAsnGlnGlu	160	
QY 564	CTGCATCAGTCAACACCGGGGATAAATCTGGATTGGTTCCGCGCTCAAGGTGAAGAT	623	
DB 161	-----ProSerAspSerAlaGlyGluGlnGluThrGlnProGluGlu	175	
QY 624	AATACCTAAAGAGACACTGTAATATCCAGACAGCTGAAGAGCAACCAAA	678	
DB 176	GlyAsn-ValGluSerProProSerSerProGluAsnSerGlnGluGlnProGln	193	
RESULT 16			
ID Q99MM1	PRELIMINARY;	PRT:	413 AA.
AC Q99MM1	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT 01-JUN-2001	(TrEMBLrel. 17, Last sequence update)		
DT 01-DEC-2001	(TrEMBLrel. 19, Last annotation update)		
DE WSL-1-LIKE PROTEIN.			
GN TNFRSF12.			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX NCBI_TaxID=10090;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=129/SV;			
RX MEDLINE=21158384; PubMed=11261933;			
RA Wang E.C.Y., Kitson J., Thern A., Williamson J., Parrow S.N.,			
RA Owen M.J.;			
RT "Genomic structure, expression, and chromosome mapping of the mouse			
RT homologue for the WSL-1 (DR3, Apo3, TRAMP, LARD, TR3, TNFRSF12)			
RT gene.;			
RL Immunogenetics 53:59-63(2001).			
DR EMBL; AF329969; AAK11256.1; -;			
DR HSP; P25942; ICDP.			
DR MGD; MGI:1934667; Tnfrsf12.			
DR InterPro; IPR000488; Death.			
DR InterPro; IPR000561; EGF-like.			
DR InterPro; IPR001368; TNFR_c6.			
DR Pfam; PF00020; TNFR_c6; 3.			
DR SMART; SM00005; DEATH; 1.			
DR SMART; SM00208; TNFR; 3.			
DR PROSITE; PS00017; DEATH_DOMAIN; 1.			
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.			
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.			
DR PROSITE; PS50050; TNFR_NGFR_2; 1.			
SQ SEQUENCE 413 AA; 44453 MW; 69F21B85D0DABABF CRC64;			
Alignment Scores:			
Pred. No.:	0.209	Length:	413
US-09-658-824-808 (1-781) x Q99MM1 (1-413)			
QY 677	TGTGTGCTCTTTCACCTGCTTTCAGCATTTTACAGTGTCTCTTTAGGTATTATCTTC	618	
DB 137	CysGlyLysSerSerProPheSer-----CysValProCysGly-AlaThrTh	152	
QY 617	ACCTTGACCCCGGAACCAATCCA--GATTTATCCCGGTGTT-----GA	573	
DB 152	rProValHisGluAlaProThrProArgProCysLeuProGlyPheTyrIleArgGlyAs	172	
QY 572	CTGATGACGCTCTTCAGATCACCTTCATGTCGCGCACTGGGATCTCAGTGTATCTCG	513	
DB 172	nAspCysThrSerCys-----ProThrGlyPheSerSerVal-----	184	
QY 512	ATCTTCTTCTGCTGCTGCCAGGTGTAGGATCCGACTTTTCAGCTGCTGCTTCTTCTT	453	
DB 185	-----CysProLysAlaCysThrAlaValCysGlyTrpLysGlnMetPh	199	
QY 452	TGGGGCTCTCCATTACTGCTGGCTGCTGGGACTGGTGTGTGTCTGTCTTGTGTGTCT	393	
DB 199	eTrpValGlnValLeuLeuGlyVal---AlaPheLeuPheGlyAlaIleLeuIleCysAl	218	
QY 392	CGCTCATGCCCTCTCTCTGCTGCTGCTGCGAGTCCAGATCTTCTGTGGGGTTTC	333	
DB 218	a-----TyrCysArgTrpG1	223	
QY 332	CAGGTGC-----TTCTGGGACTCGGATACCCCAACAGCACCCCA---GCTTCGCTCA	282	
DB 223	nProCysLysAlaValValThrAlaAspThrAlaGlyThrGluProLeuAlaSerProG1	243	
QY 281	GCCCTCCACTCTCTCGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	222	
DB 243	nThrAlaHisLeuSerAla-----	249	
QY 221	CCATGGGCGTGGCACCCGATCAGAGGGACGAGCTTCGAGGGCTTCTCACTCTGATCTC	162	
DB 250	-----SerAspSe	252	
QY 161	ACTGACCACACACCCCTCCGGAAGCTGCTGGTTCCTCTCTCACTCTCACTCACACTCAA	102	
DB 252	rAla--HisThrLeuLeuAlaProProSerSerThrGlyLysIleCysThrThrValGln	271	
QY 101	CTCCAGTTGGATCT-----	78	
DB 272	Leu---ValGlyAsnAsnTrpThrProGlyLeuSerGlnThrGlnGluValValCysGly	290	
QY 77	-----CCTACCGCGCTGCTCAGTAGCGGAGAAAGAA	45	
DB 291	GlnAlaSerGlnProTrpAspGlnLeuProAsnArg-----ThrLeuGlyThrProLeu	308	
QY 44	TCCAGACTCAGGACCGGAGTCCCGGCTCCAGCTCCGCGCC	3	
DB 309	AlaSerProLeuSerProAlaProProAlaGlySerProAla	322	
RESULT 17			
ID Q9U229	PRELIMINARY;	PRT:	643 AA.
AC Q9U229	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT 01-MAY-2000	(TrEMBLrel. 13, Last sequence update)		
DT 01-DEC-2001	(TrEMBLrel. 19, Last annotation update)		
DE Y56A3A.32 PROTEIN.			
GN Y56A3A.32			
OS Caenorhabditis elegans.			
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;			
OC Rhabditidae; Peloderinae; Caenorhabditis.			
OX NCBI_TaxID=6239;			

OS	Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).	
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;	
OC	Gammaherpesvirinae; Rhadinovirus.	
OX	NCBI_TaxID=37296;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21165304; PubMed=11264383;	
RA	Pilot T., Tramlair M., Coppey M., Nicolas J.C., Marechal V.;	
RT	*Close but distinct regions of human herpesvirus 8 latency-associated	
RT	nuclear antigen 1 are responsible for nuclear targeting and binding to	
RT	human mitotic chromosomes.*;	
RL	J. Virol. 75:3948-3959(2001).	
DR	EMBL; AF305694; AAG27458.1; -.	
DR	HSP; P02649; INFN.	
DR	InterPro; IPR002965; P_rich_extensn.	
DR	InterPro; IPR002017; Spectrin.	
DR	InterPro; IPR000533; Tropomyosin.	
DR	PRINTS; PR01217; PRICHEXTENS.	
DR	PRINTS; PR00194; TROPOMYOSIN.	
FT	NON_TER 1036 1036	
SQ	SEQUENCE 1036 AA; 119328 MW; 2959EDD2C1C042B8 CRC64;	
Alignment Scores:		
Pred. No.:	0.269	Length: 1036
Score:	102.50	Matches: 33
Percent Similarity:	37.76%	Conservative: 21
Best Local Similarity:	23.08%	Mismatches: 43
Query Match:	7.43%	Indels: 46
DB:	12	Gaps: 4
US-09-658-824-808 (1-781) x Q9DUM3 (1-1036)		
QY	261	AGGGCCGAGAGTGGAGGGCTCAGCGAAGCTGGGGTGTGTTGGGGTATCCGAGTCC 320
Db	432	LysLysGluAspGluGluAspGlyGlyAspGlyAsnLysThrLeuSerIleGlnSerSer 451
QY	321	CAGAAGCACTTGGAAACCCCGACAGAAAGATTCTGGACTCCCGACAGCGGACGAGGAGGG 380
Db	452	GlnGlnGlnGlnGluProGlnGlnGln----- 460
QY	381	ACGGCATGAGCGACACACACAAACACAGAACCCACACAGC-----CAGTCCCGAGGAG 431
Db	461	-----GluProGlnGlnGlnGluProGlnGlnGlnGlu 471
QY	432	CCCAAGTGGAGAGCCCAAGAAAGAACACAGAGCTGAAAGTCGGATCCCTACACCT 491
Db	472	ProGlnGlnGlnGluProGlnGlnGlnGluProLeuGlnGluProGlnGlnGlnGluPro 491
QY	492	GGCAGCAGACAGAAAGATCAGGATACAGTCCAGTCCAGTCCCGACATGGAAGGT 551
Db	492	GlnGlnGlnGluProGlnGlnGlnGluProGlnGln-----GlnGlu 505
QY	552	GATCTGCAAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGGATTGGGTTCCGCGGT 611
Db	506	ProLeuGlnGluProGlnGln-GlnGlu----- 514
QY	612	CAAGGTGAAGATAATACCTAAAGAGAAACACTGTAAATGCCAGAAGAGGTGAAGAGCA 671
Db	515	-----ProGlnGlnGlnGluProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 529
QY	672	ACCACAA 678
Db	529	uProGln 531
RESULT 19		
Q98148		
ID	Q98148	PRELIMINARY; PRT; 1162 AA.
AC	Q98148;	
DT	01-FEB-1997 (TrEMBLrel. 02, Created)	
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE	ORF73 HOM <sup>1</sup> 15.	
OS	Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).	



Db	1346	ProLeuAlaProLeu-----	:::	::   ::	--ProValLeu-AlaProSerProGl	1358
Qy	418	CTGTGTTG---GTTCTGTGGTTCGCCTCATCGCCGCCCTCCTCCCTGGTGCCCGCTCT		::: ::   :::		362
Db	1358	yAlaAlaProValLeuAlaSerSerGlnThrProValProValMet-----		: : :		1373
Qy	361	GCGAGACTCCAAGATCTCTGTGCGGGTTCACGGTGCTCTGGGACTCGGATACCCECAAC			:	302
Db	1374	----AlaProSerSerThrProGlyThrSerLeuAlaSerAlaSerProValPro----			:	1390
Qy	301	AGCACCCCAGCTTCGCGTCGAGGCCCTCCACT--			-CCTCGGCC	263
Db	1391	-AlapProThrProValLeuAlaProSerSerThrGlnThrMetLeuProAlaProValPr				1410
Qy	262	CTCCTTCCTCCCTAGTTATGGCGGTGACGACAGACAAGAAGTCCATGGCGGTGGCACCGCA				203
Db	1410	oSerProLeuProSerProAlaSerThrGlnThr-----		:	--LeuAlaLeuAlaPro-A	1427
Qy	202	TGAGAGGAGCAGCGACTTCGAGGGCTTCTCACTCTGAGTCACTG- - - - -			-----ACCACA	152
Db	1427	laLeuAlaProThr-LeuGlyGlySerProSerProGlnThrLeuSerLeuGlyThrGly				1446
Qy	151	CAAACCTCCGGGAAGCCTGCTGCTCTCTCACTCTCACTC-				111
Db	1447	AsnProGlnGly--ProPheProThrGlnThrLeuSerLeuThrProAlaSerSerLeuV				1466
Qy	110	-----ACACTTCAACTCCCCAGTTGGATCTGCCTGTGGGACCTACCCGCC				68
Db	1466	alProThrProAlaGlnThrLeuSerLeuAlaProGlyProProLeuGlyProThrGInt				1486
Qy	67	GTGCTCAGTAGGAGGAG-----AAAGAATTCCAGACCTCAGGGACCCGAGTCGC				20
Db	1486	hrLeuSerLeuAlaProAlaProProLeuAlaProAlaSerProValGlyPro--Alap				1505
Qy	19	CGGCTCAGCAGC 9				
Db	1505	roAlaHlsThr 1508				
RESULT	21					
Q33572		PRELIMINARY; PRT;		350 AA.		
ID	AC	Q33572				
DT	01-NOV-1996	(TEMBLrel. 01, Created)				
DT	01-NOV-1996	(TEMBLrel. 01, Last sequence update)				
DE	01-DEC-2001	(TEMBLrel. 19, Last annotation update)				
DE	KINETOPLAST APOCYTOCHROME B.					
GN	COB OR CYTB.					
OS	Trypanosoma brucei.					
OG	Mitochondrion.					
OC	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.					
OX	NCBITaxID=5691;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
EX	MEDLINE=85086128; PubMed=6513991;					
RA	Johnson B.J.B., Hill G.C., Doneison J.E.; .					
FT	"The maxicircle of <i>Trypanosoma brucei</i> kinetoplast DNA encodes apocytochrome b.";					
RL	Mol. Biochem. Parasitol. 13:135-146(1984).					
CC	-I- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL COUPLED TO ATP SYNTHESIS (BY SIMILARITY).					
CC	-I- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN (BY SIMILARITY).					
CC	-I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B6-F ARE: CYTOCHROME B6, 17 KDA POLYPEPTIDE (PETD), CYTOCHROME F AND THE RIESKE PROTEIN (BY SIMILARITY).					
CC	-I- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.					
DR	EMBL; M17998; AAA32115.1;- ;					
DR	InterPro; IPR000179; Cyt_b.b6.					
DR	Pfam; PF00033; cytochrome_b_c_1.					
DR	Pfam; PF00033; cytochrome_b_n; 1.					

D	b	646	userAlaGlyValSerProLysArgArg-----AlaLeuProServAlA]	6																	
Q	y	163	TCA-CTGACCACAAACCCCTCGGAAGCCTGCTGGTTCTTCTTCACCTCTCACACTCACACATT	105																	
D	b	661	aGlyLeuPheThrGlnProSerGlySer-----ProLeuAsnLeuAsnSer---Me	677																	
Q	y	104	CAACTCCAGTGTGATCTGCTGTGGACCTTACC CGCGGTGTCT	62																	
D	b	677	tValSerAsnIleAsnGlnProLeuGluIleThrAlaIleSer	691																	
<b>RESULT 23</b>																					
Q	SQN08		PRELIMINARY;	PRT; 1054 AA.																	
A	C	QSN08:																			
D	T	01-MAY-2000 (TrEMBLrel. 13, Created)																			
D	T	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)																			
D	T	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)																			
D	E	HYPOTHETICAL PROTEIN.																			
O	S	Oryza sativa (Rice).																			
O	C	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;																			
O	C	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;																			
O	C	Eriarthoidae; Oryzeae; Oryza.																			
O	X	NCBI_TaxId=4530;																			
R	N	[1]																			
R	P	SEQUENCE FROM N.A.																			
R	C	STRAIN-CV. NIPPONBARE;																			
R	A	Sasaki T., Matsumoto T., Yamamoto K.;																			
R	T	*Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC																			
R	T	clone:P0335G04.*;																			
R	L	Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.																			
R	L	EMBL; AP000399; BAA83569.1; -																			
K	W	Hypothetical protein.																			
S	Q	SEQUENCE 1054 AA; 113435 MW; 97DD6F2144766422 CRC64;																			
<b>Alignment Scores:</b>																					
P	r	e	d.	N	o.	:	0.421	L	e	n	g	t	h.	1054							
S	c	o	r	e.			100.50	M	a	t	c	h	e	s.	53						
P	e	r	c	e	n	t	32.06%	C	o	n	s	e	r	v	a	t	i	v	e	.	14
B	e	s	t	L	o	c	a	l	S	i	m	i	l	a	r	i	t	y.	78		
Q	u	e	r	y	M	a	t	c	h.	7.28%	I	n	d	e	l	s.	64				
D	B	:	10	G	a	p	s.	7													
<b>US-09-658-824-808 (1-781) x Q9SN08 (1-1054)</b>																					
Q	y	123	AGAGAACCAGCAGCGCTTCGAGGAGTGTTGTGTCAGTCACTCAGAGTGAAGGCCCT	18																	
D	b	321	ArgglyGluGlnSerThrProArgGlyGlyArgAlaSerGlySerArgAspArgGlyPro	340																	
Q	y	183	CGAAGTCGTGCTCCTCTCATGCG-	212																	
D	b	341	GlySerSerArgProAlaAspAlaArgGlyLysArgLysGlnGlyGlyThrProProPro	360																	
Q	y	213	CGCCCATGGACCTTCTTGCTCTGCACGGCCATACTAGGAGGAGAGGAGCGCGAGGAG	272																	
D	b	361	SerPro-----ProArgGlyGly-----GlyAlaAlaArgAlaSerAsn	373																	
Q	y	273	TGGAGGGCTCAGCGAAGCTGGGTGCTCTTGGGGGTATCCGAGTCCCAGAACCACTG	332																	
D	b	374	ArgArgProGluGlyAlaAlaProThr-----	382																	
Q	y	333	GAACCCGCACAGAGATTCTGGACTCCCCAGAGGGACACAGGAGGACGGCATGAGCG	392																	
D	b	383	-----SerGlnProGluGlyGluArgLysLysLysArgProArg	395																	
Q	y	393	ACACACAACACAGAACCAACAC-	437																	
D	b	396	LysMetGlyGluThrGluProSerArgGlyAsnLeuIleSerProProLysTrpSerPhe	415																	
Q	y	438	AATGGAGCCCCAAAAGAAGAACACAGCAGCTGAAGTCGGGATCTCTACACCTGGCCAG	497																	
D	b	416	AsnArgProProArgSerGluIleProSerArgProSerArgHisSerLysSerGlyGln	435																	
Q	y	498	CAGACAGAAGAAGATCAGGATACAGCTGAGATC-----	530																	

```
Db 436 SerGluAlaGluAspProAlaAlaAlaGluAlaLeuArgArgGluSerAspArgArgGlu 455
QY 531 -----CCAGTGGCGGACATGGAAGTGCATCTCGAAGAGTCAGTCAGTCAGTCAAAACACC 581
Db 456 AlaAlaAspArgLeuArgGluAlaGluGluThrAlaGlnGluAlaAlaArgAla----- 473
QY 582 GGGGATAAATCTGGATTCTGGTTCGGCTCGCGCTCAAGGTGAAGATAAATACCTAAAGAGGAACA 641
Db 474 -----ArgGlnAlaGluGluThrAla-ArgGluGluAl 484
QY 642 CTGTAATAATCCAGACAGCAGTCAA 666
Db 484 aAlaArgAlaArgGlnAlaGlyGlu 492
RESULT 24
Q27033
ID Q27033 PRELIMINARY; PRT; 480 AA.
AC Q27033;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE MEMBRANE PROTEIN.
OS Theileria parva.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC Theileria.
OX NCBI_TaxID=5875;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94088665; PubMed=8264721;
RA Baylis H.A., Allsopp B.A., Hall R., Carrington M.;
RT "Characterisation of a glutamine- and proline-rich protein (QP
RL Mol. Biochem. Parasitol. 61:171-178(1993).
DR EMBL; L06323; AAA18800.1; -.
SQ SEQUENCE 480 AA; 52388 MW; 619B46AE128EA2E3 CRC64;
Alignment Scores:
Pred. No.: 0.417 Length: 480
Score: 100.00 Matches: 40
Percent Similarity: 43.66% Conservative: 22
Best Local Similarity: 28.17% Mismatches: 42
Query Match: 7.25% Indels: 38
DB: 5 Gaps: 8
US-09-658-824-808 (1-781) x Q27033 (1-480)
QY 341 ACAGAAAGATTCTGGACTCCCGACGGGACAGGAGGAGGCGGATGAGCGACACACAC 400
Db 68 ThrSerGlnGlnGlyLysProGlnProAspGlnProGlnAsp- --GlnProAspGlnHls 86
QY 401 AAACAC-----AGAACCCACAGCCAGTC 424
Db 87 GlnGlnProThrGlnGlyAspThrSerGlyGlnGlnGlyProAspThrProGlnProIle 106
QY 425 CCAGGAGCCAGTAATGTGAGAGCCCAAAAGAA-----GAACC 463
Db 107 -GlnGluProSer---GlyProValGlnProAspGlnThrGlyGlnGlyProValGluPr 125
QY 464 AGCAGCTGAATCGGGATCCCT-----ACACTGGCGGACGACAGAGAAGA 511
Db 125 oValAspGlnGlnGlnProThrGlnGlyAspThrSerGlyGlnGlnGlnGlnPr 145
QY 512 TCAGGATACAGTCGATCCCGGCGGACATGGAAGGTGATCTGCAAGAGCTGCATCA 571
Db 145 oGlnAspGln-----ProValGlnGlnAspGlyGlnAspSerGlnGlyThrPr 162
QY 572 GTCAACACCGGGGATAAATCTGGATTCTGGTTCGGGGTCAAGGTGAAGATAATACCT- 630
Db 162 oGluGlnThrProAspGlnSerGly-----GlnGlnProGlyProAspThr-PROA 179
QY 631 -----AAAGAGGAACACTGTAAATCCCAAGACGAGGTGAAGAGCAACACC 676
```

```
Db 179 spGlnProValTyrGlnGlnGlnProValGlnGlnProSerGlyGlnGlnGlnProG 199
QY 677 AA 678
Db 199 ln 199
RESULT 25
Q94E85
ID Q94E85 PRELIMINARY; PRT; 181 AA.
AC Q94E85;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE B1045D11.17 PROTEIN.
GN B1045D11.17.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
clone:B1045D11.17."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003199; BAB61149.1; -.
SQ SEQUENCE 181 AA; 19391 MW; 4B7CDC5D082442AF CRC64;
Alignment Scores:
Pred. No.: 0.401 Length: 181
Score: 99.50 Matches: 49
Percent Similarity: 43.68% Conservative: 27
Best Local Similarity: 28.16% Mismatches: 58
Query Match: 7.21% Indels: 40
DB: 10 Gaps: 10
US-09-658-824-808 (1-781) x Q94E85 (1-181)
QY 164 CTCAGAGTCAGAAAGCTCGAAGTCGTCCTCTCA-----TGGCGTGCCACG 214
Db 7 lleArgTTPArgArgProAlaAlaLeuSerProLeuSerProAlaGlySerGlySerPro 26
QY 215 OCCATGGACCTTCTGTCGTCAGGCCATAACTAGGAGGAGGAGGAGGAGGAGGAGTG 274
Db 27 ArgLeuLeuPheLeuProSerProAspPro-----AlaGluGlyArgGlyVal 42
QY 275 GAGGGCTCAGCGGAAGCTGGGTGCTGTT-----GGGGGTATCGGAGTCCCA 322
Db 43 AspGly-----GlyAlaValGluGlyGluProGlyGlyGlnArgGluHls 57
QY 323 GAAGCACCTTGA-----ACCCGACAGAAAGAT-----TCTGGACTCCCCAGA 364
Db 58 GlySerGlyGlyArgProLeuSerProThrArgAspArgAlaSerAlaProLeuProSer 77
QY 365 CGGACACAGGAGAGCGGATGAGCGACACACACACACACACACACACACACACACATC 424
Db 78 ProProLeuProArgSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 95
QY 425 CCAGGAGCCACAG---TAATGAGAGCCCAAAAGAAAGAACCA----- 464
Db 96 ArgGlyAlaArgSerArgTrpArgAlaValArgArg-ProProProThrMetThrArgTh 115
QY 465 -----GCAGCTAAAGTCGGGATCTTACCTCGGCGAGCAGACAGACAGAAAGATCAGGA 517
Db 115 rGlyValAlaSerSerAlaLysAspGluAspAspGluGluGluAspAspLysGluGlu 135
QY 518 TACAGCTCAGATCCCGAGTCGCGGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAA 577
Db 135 uAlaAlaAlaProProAsnSerArgArg-----TyrGluSerAl 149
```



[illegible]

```
Q99JV6
ID Q99JV6 PRELIMINARY; PRT; 630 AA.
AC Q99JV6
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 67.2 KDA PROTEIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005641; AA005641.1;
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 630 AA; 67197 MW; 58DD514CADBAE492 CRC64;

Alignment Scores:
Pred. No.: 0.543 Length: 630
Score: 99.00 Matches: 44
Percent Similarity: 37.14% Conservative: 21
Best Local Similarity: 25.14% Mismatches: 70
Query Match: 7.17% Indels: 40
DB: 11 Gaps: 9

US-09-658-824-808 (1-781) x Q99JV6 (1-630)
QY 138 CTTCCGGAGGTTGTGGTCAGTCAC-----TCAGAGTGAGAGGCCCTCGAAGTCGT 191
Db 49 ILeProHisAspLysValValAspGluGlnAspGluGlyProGluGluLys 68
QY 192 CGTCCTCTCATCGGGTCCACGCCCTGACCTTCTGTCTGTCACGCCATAACTAG 251
Db 69 LysProProArgLeuProAspGlu----- 77
QY 252 GGAGGAGGAGGCGGAGGAGTGGAGGGCTCAGCGCAAGCTGGGTGCTGTGGGGGTA 311
Db 78 --GlyAspProAla-----GlyArgGlyGlnGlyAlaProProLeuProGlu 92
QY 312 TCCAGAGTCCCAAGACACCTGGAACCCGACAGAGAGTCTGGACTCCCGAGCGGAC 371
Db 93 SerGluLysGluLysGln--GluProGluArg-----GlyGlyGluGlyLys 107
QY 372 AGGAGAGGAGCGGATGAGCGACA-----CACACAACACAGAACCCACAGC 419
Db 108 ArgProGlyGlnValLeuAlaValGlyGluThrgluHisProGlnLysValProGluAla 127
QY 420 CAGTCCCGAGGAGCAGTAAATGGAGAGCCCAAAAGAACAGCAGCAGCTGAAGTCGG 479
Db 128 AsnGlyGlnProProVal-----GlnProArgLysGluAspSerArgProGlyAsnArg 145
QY 480 GATCCTACACTGGCGACAGACAGAGAAGAT-----CAGGAT 518
Db 146 AspProGlnProAlaAlaGlnAlaArgAspSerValGluLeuLysAlaLeuAlaAsp 165
QY 519 ACAGCTGAGATCCCGAGCGGACATGGAAGGTGATCTG-----CAAGAGCTG 566
Db 166 AspGlyArgGluProAlaGlnGlnLysAlaGlyAlaLeuTrpLysProValGluSerAla 185
QY 567 CATCAGTCAACACCGGGGATAAATCTGGATTTGGTTCGGCGGT 611
Db 186 AlaGluSerAspAlaGlyGlyLysAlaGlyLeuProValGlnArg 200

RESULT 29
Q76871
ID Q76871 PRELIMINARY; PRT; 773 AA.
AC Q76871;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE EG:100G7.2 PROTEIN.
GN CG3591 OR EG:100G7.2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Murphy L., Harris D., Barrell B.;
RL *sequencing the distal x chromosome of Drosophila melanogaster.*;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031366; CAA20518.1;
DR FlyBase; FBgn0014097; CG3591.
SQ SEQUENCE 773 AA; 86307 MW; 1D073855D5E7A983 CRC64;

Alignment Scores:
Pred. No.: 0.561 Length: 773
Score: 99.00 Matches: 34
Percent Similarity: 42.74% Conservative: 19
Best Local Similarity: 27.42% Mismatches: 54
Query Match: 7.17% Indels: 17
DB: 5 Gaps: 4

US-09-658-824-808 (1-781) x Q76871 (1-773)
QY 221 GACCTTCTGTCTGTACGGCCATACTAGGAGGAGGAGGCGCGAGGTGAGGGG 280
Db 658 AspLeuLeuThrglnAsnArgArgPheAlaLeuArgLysThrglyArgSerLeuGluLe 677
QY 281 CTCAGCGAAGCTGGGTGCTGTGGGGTATCCGAGTCCCAAGACACTGGAAACCCG 340
Db 678 ProGluGlnValAlaProLysGlyAspLeuGlnLysGluGluLysProLysGluGlu 697
QY 341 ACAGAAGATCTCGATCCCGAGCGGACCGAGGAGGAGGCGCGATGAGCGACACAC 400
Db 698 GluGlnLysGluLysLeuProLysGluGluValGln-----LeuGluGluLeLys 714
QY 401 AAACACAGAACCCACACAGC-----CAGTCCCGAGGAGCCCAAGTAAATGGAGCGCCCA 451
Db 715 LysGlu-GluProGlnLysGluGluLeuGlnLysGluGluProGlnLysGluGluProG 734
QY 452 AAAAGAAGAACCCACAGCAGCTGAA-----AGTCGGGATCTCTACACCTGGCGAGCAGACA 505
Db 734 nLysGluGluProArgLysGluGluProGlnLysGluGluProGlnLysGluGluProG 754
QY 506 AGAAGATCAGGATACAGCTGAGATCCCGAGTCCCGAGTGGAGGAGGTGATCTGCAAGAGCT 565
Db 754 nLysGluGluProLysValGluThrPro-----GlnProLe 766
QY 566 GCATCAGTCA 575
Db 766 uGluGlnSer 769

RESULT 30
Q9W4V0
ID Q9W4V0 PRELIMINARY; PRT; 773 AA.
AC Q9W4V0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ANON-3CB PROTEIN.
GN CG3591 OR ANON-3CB OR EG:100G7.2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
```

[illegible]

Percent Similarity:	42.74%	Conservative:	19
Best Local Similarity:	27.42%	Mismatches:	54
Query Match:	7.17%	Indels:	17
DB:	5	Gaps:	4

US-09 658-824-808 (1-781) x Q9I7X0 (1-821)

QY	221	GACCTCTCTCTCTGCTACCGCCATACCTAGGAGGAGGAGGCCGAGAGTGGAGGG	280
Db	706	AspLeuLeuThrGlnAsnArgArgPheAlaLeuArgLysThrGlyArgSerLeuGluIle	725
QY	281	CTCAGCGCAAGCTGGGTGCTGTTGGGGTATCCGAGTCCCAGAACCACTGGAACCCCG	340
Db	726	ProGluGluGlnValAlaProLysGlyAspLeuGlnLysGluGluLysProLysGluGlu	745
QY	341	ACAGAAGATTCTGCATCCCGACAGCGGACACGAGGAGCGCATGAGCCACACACAC	400
Db	746	GluGlnLysGluLysLeuProLysGluGluValGln-----LeuGluGluIleLys	762
QY	401	AAACACAGAACCACACAGC-----CAGTCCCAGAGCCCGCAATGAGAGCCCCA	451
Db	763	LysGlu-GluProGlnLysGluGluLeuGlnLysGluGluProGlnLysGluGluProGln	782
QY	452	AAAGAAGAACCAAGCAGCTGAA-----AGTCGGGATCCTACACCTGGCGCACACAGA	505
Db	782	nLysGluGluProArgLysGluGluProGlnLysGluGluProGlnLysGluGluProGln	802
QY	506	AGAAGATCATCAGGATACAGCTCAGATCCAGTCGCGCAGCATGGAAGGTGCTCATC	565
Db	802	nLysGluGluProLysValGluThrPro-----GlnProLeu	814
QY	566	GCATCAGTCA	575
Db	814	uGluGlnSer	817

RESULT 32

Q41516	PRELIMINARY;	PRT;	405 AA.
ID	Q41516		
AC	Q41516;		
QC	Q41516;		
DT	01-NOV-1996 (TReMBLrel. 01, Created)		
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)		
DE	HIGH-MOLECULAR-WEIGHT GLUTENIN.		
DE	Triticum aestivum (Wheat).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;		
OC	Trifurcaceae; Triticum.		
OC	NCBI_TaxID=4565;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=ASARCE; TISSUE=ENDOSPERM;		
RA	Xin H., Jiang Y., Huang W., Gu Q., Sun C.;		
RC	"Nucleotide sequence of a high-molecular-weight glutenin gene in		
RT	wheat.;"		
RT	Acta Bot. Sin. 34:729-735(1992).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=ASARCE; TISSUE=ENDOSPERM;		
RA	Sun C.;		
RC	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.		
RL	EMBL; U019774; AAA62315.1; -.		
DR	SEQUENCE 405 AA; 44342 MW; CEAA1861335C5736 CRC64;		
SQ			

Alignment Scores:	
Pred. No.:	0.568
Score:	98.50
Percent Similarity:	32.08%
Best Local Similarity:	21.70%
Query Match:	7.14%
DB:	10
	6
Length:	405
Matches:	46
Conservative:	22
Mismatches:	73
Indels:	72
Gaps:	6

US-09-658-824-808 (1-781) x Q41516 (1-405)

QY	61	GAGACACGGCGGTAGTTCACAGCAGCATCCAACTGGGAGTTGAAGTGTCAGTGCAGT	120
Db	50	GlnLeuArgAspValSerProGluCysGlnProValGlyGlyGlyProValAlaAraGln	69
QY	121	GAAGAGGAACCAAGCAGGCTTCCGGGAGGCTTGTGTGGTCAGTGACTCAGATGAGAAGGCC	180
Db	70	TyrGluGln-----	72
QY	181	CTCGAAGTCGTGTCCTCTCATCGGTGCCACGCCCATGGACCTTCTCTCGTCACG	240
Db	73	---GlnValValPro-----	77
QY	241	GCCATAACTAGGAGGAGGAGGCCGAGGAGTAGTGGAGG-----	279
Db	78	-----proSlyGlyIleSperPheTyrProGlyGluThr	89
QY	280	-----GCTCAGCGCAAGCTGGGCTGCTGTGGGGGTATCCGAGT-----	318
Db	90	ProProGlnGlnLeuGlnGlnSerIleLeuTrpGlyIleProAlaLeuLeuArgArgTyr	109
QY	319	CCCAGAAGCAGCTGGAACCCGACAGAGATTCTGGAGATCCCCAGAGGGGA-----	369
Db	110	TyrLeuSerValThrSerProGlnArgValSerTyrTyrProGlyGlnAlaSerSerGln	129
QY	370	-----CCAGGAGAGGCGGCGATGATGAGCGACGACAC-ACAACACACA	407
Db	130	ArgProGlyGlnGlyGlnGlnProGlyGlnGlyGlnGlnGlnTyrTyrLeuThrSerPro	149
QY	408	GAACCACACAGCCAGTCCAGGAGCCCGCAGTAATGGAGAG-----	446
Db	150	GlnGlnSerGlyGlnTrpGlnGlnProGlyGlnGlnGlnSerGlyTyrTyrProThrSer	169
QY	447	CCCCAAAAGAAGAACACAGCAGCTGAAAGTCGGGATCCTACACCTGGGAGCAGACAGAA	506
Db	170	ProGlnGlnSerGlyGlnGlnGlnProGlyTyrTyrProThrSerProTrpGlnProGlu	189
QY	507	GAAGATCAGGATACAGCTGAGATCCCAGTCCGCGCACATGGAAGCGATCTGCAAGAGCTG	566
Db	190	GlnLeuGlnProThrGlnGlnGlnArgGlnGlnProGlyGlnGlyGlnGlnGln	209
QY	567	CATCAGTCAAACACCGGGGATAAATCTGGATTGGG	602
Db	210	ArgGlnGlyGlnGlnGlnGlnSerGlyGlnGly	221
RESULT	33		
ID	Q9SYVO	PRELIMINARY;	PRT; 405 AA.
AC	Q9SYVO;		
DC	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DI	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	PUTATIVE HIGH MOLECULAR WEIGHT GLUTENIN PRECURSOR.		
GLU-A1.			
OS	Triticum aestivum (Wheat).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;		
OC	Triticeae; Triticum.		
RN	NCBI_TaxID=4565;		
FX	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=CV, PANE-247;		
RA	De Bustos A., Rubio P., Jouve N.;		
RT	"Molecular characterisation of the inactive allele of the gene Glu-A1		
RT	and the development of a set of AS-PCR markers for HMW glutenins of		
RT	wheat.";		
RL	Theor. Appl. Genet. 100:1085-1094(2000).		
DR	EMBL; AF145590; AAD32223.1; -		
DR	InterPro; IPR001419; Glutenin.		
DR	PRINTS; PR00210; GLUTENIN.		
KW	Signal.		
FT	SIGNAL	1	21 POTENTIAL.
FT	CHAIN	22	405 PUTATIVE HIGH MOLECULAR WEIGHT GLUTENIN.
SQ	SEQUENCE	405 AA,	44396 MW; 3D74186D86538D44 CRC64:

Alignment Scores:  
 Pred. No.: 0.568 Length: 405  
 Score: 98.50 Matches: 46  
 Percent Similarity: 32.08% Conservative: 22  
 Best Local Similarity: 21.70% Mismatches: 73  
 Query Match: 7.14% Indels: 72  
 DB: 10 Gaps: 6

US-09-658-824-808 (1-781) x Q95Y0 (1-405)

QY 61 GAGACACGGGGTGGTCCAGGAGGAGTCCAACTGGAGTTGAAGTGTGAGTGAGAGT 120  
 Db 50 GlnLeuArgAspValSerProGluGlyGlnProValGlnGlyGlyProValAlaArgGln 69  
 QY 121 GAAGAGAAACAGCAGGCTCCGGAGGGTGTGTGTCAGTCACTCAGAGTGAGAGGCC 180  
 Db 70 TyrGluGln----- 72

QY 181 CTCGAAGTCGTCTCCCTCTCATCGGTGCCAGGCCCATGGACCTTCTTGTCTCTGCACG 240  
 Db 73 ---GlnValValPro----- 77

QY 241 GCATAACTAGGAGGAGGAGGCGGAGAGTGGAGG----- 279  
 Db 78 -----ProLysGlyGlySerPheTyrProGlyGluThr 89

QY 280 -----GCTCAGCGGAAGCTGGGTCTGTGGGGTATCCGAGT----- 318  
 Db 90 ProProGlnGlnLeuGlnGlnSerIleLeuTrpGlyIleProAlaLeuArgTyr 109

QY 319 CCAGAACACCTGGAAACCCGACAGAGATTCTGGACTCCCGACAGCGGA----- 369  
 Db 110 TyrLeuSerValThrSerProGlnArgValSerTyrTyrProGlyGlnAlaSerSerGln 129

QY 370 -----CCAGAGAGGACGCATGACGACACAC-ACAAACACA 407  
 Db 130 ArgProGlyGlnGlnGlnProGlyGlnGlnGlnGlnGlnTyrTyrLeuThrSerPro 149

QY 408 GAACACACAGCCAGTCCAGGAGGCCAGTATGGAG----- 446  
 Db 150 GlnGlnSerGlyGlnTrpGlnGlnProGlyGlnGlyGlnSerGlyTyrTyrProThrSer 169

QY 447 CCCCCAAAAGAAACACAGCAGCTGAACTGGGATCTACACCTGGCGAGCAGACAGAA 506  
 Db 170 ProGlnGlnSerGlyGlnGlnGlnProGlyTyrTyrProThrSerProTrpGlnProGlu 189

QY 507 GAAGATCAGGATACAGTCCAGTCCAGTGGCGGACATGGAAGGTGATCTGCAAGAGCTG 566  
 Db 190 GlnLeuGlnGlnProThrGlnGlnGlnArgGlnGlnProGlyGlnGlyGlnGlnLeu 209

QY 567 CATCAGTCAACACCGGGGATAAATCTGGATTGGG 602  
 Db 210 ArgGlnGlyGlnGlnGlnSerGlyGlnGly 221

RESULT 34  
 Q95KP7  
 ID Q95KP7 PRELIMINARY; PRT; 847 AA.  
 AC Q95KP7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE TREACLE (FRAGMENT).  
 GN TCOF1.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Haworth K.E.;  
 RT "Canine TCOF1: Cloning, chromosome assignment and genetic analysis in  
 dogs with different face and head types.\*";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ302713; CAC88122.1; JOINED.  
 DR EMBL; AJ302714; CAC88122.1; JOINED.  
 DR EMBL; AJ302715; CAC88122.1; JOINED.  
 DR EMBL; AJ302716; CAC88122.1; JOINED.  
 DR EMBL; AJ302717; CAC88122.1; JOINED.  
 DR EMBL; AJ302718; CAC88122.1; JOINED.  
 DR EMBL; AJ302719; CAC88122.1; JOINED.  
 FT NON\_TER 847  
 SQ SEQUENCE 847 AA; 85646 MW; C5A763PDF1CA2235 CRC64;

Alignment Scores:  
 Pred. No.: 0.636 Length: 847  
 Score: 98.50 Matches: 55  
 Percent Similarity: 40.10% Conservative: 22  
 Best Local Similarity: 28.65% Mismatches: 74  
 Query Match: 7.14% Indels: 41  
 DB: 6 Gaps: 9

US-09-658-824-808 (1-781) x Q95KP7 (1-847)

QY 19 GCGACTCGGGTCCCTGAGGTCTGGATTCTTCTCCGTACTGAGACACGCGGGTAGGT 78  
 Db 245 GlyLysAlaGlyPro-----SerAlaAlaGlnAlaLysThrGluLys 258

QY 79 CCACAGCAGATCCAACTGGGACTTGAAGTGTGAGTGAGTGAAGAGAACACAGGC 138  
 Db 259 ProLysGluAspSerAspSerSerGluGluAspSerGluGluProProAla 278

QY 139 TTCGAGGGTGTGTGGTCACTGAGTGTGAGAGGCCCTCGAAGTCGTCCTCCCT 198  
 Db 279 AlaLysThrProLeu---GlnValLysProSerGlyLysThrProGlnValLysAlaAla 297

QY 199 CTCATCGGTGCCACGCCCATGGACCTTCTGTCTCAGCGCCATPACTAGGGAGAA 258  
 Db 298 SerAlaSerAla---LysGluSerProArgLysGlyValProProValProGlyLys 316

QY 259 GGAGGCCGAGGACTGGAGGGCTCAGCGGAGCTGGTGTGCTGGGGGTATCCGAGT 318  
 Db 317 ValGlyProAlaAlaGly-----GlnAlaLysLys-GlyAla-----G1 329

QY 319 CCAGAGGACCTGGAAACCCGACAGAGATTCTGGACTCCCGAGC----- 365  
 Db 329 yGluGluAspProAspSerSerThrGluGluSer-AspSerGluGluGluAlaProThrA 349

QY 366 -----GGGACGAGGAGGCGCATGAGCGACACACACAAACACACAGAACACACA 417  
 Db 349 laValProProThrArg-----SerProValGlnAlaLysProSerG 363

QY 418 GCCAGTCCCGAGG-----CCCGATATGGAGAGAGAGAGAGAGAG 459  
 Db 363 yGlnAsnSerGlnValArgThrAlaSerGlyProValLysGlyProGlnLysAlaG 383

QY 460 AACACAGCAGTGAAGTCCGGATCCCTACACCTGGCGGACGACAGAGAGAGATCAGGATA 519  
 Db 383 lyProAlaAla-----ThrProValGlyLysGlnGluGluAspSerGluS 398

QY 520 CAGCTGAGATCCCGAGTCCCGACATGGAAGGT 551  
 Db 398 exSerSerGluGluGluSerAspSerGluGly 408

RESULT 35  
 Q949B8  
 ID Q949B8 PRELIMINARY; PRT; 1012 AA.  
 AC Q949B8;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 109.0 KDA PROTEIN.  
 GN C1015ERIPDK.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```
OC Ehrhartoidae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21329048; PubMed=11435398;
RA Mayer K., Murphy G., Tarchini R., Wambutt R., Volckaert G., Pohl T.,
RA Duesterhoeft A., Stiekema W., Entian K.D., Terry N., Lemcke K.,
RA Haase D., Hall C.R., van Dodeweerd A.M., Tingey S.V., Mewes H.W.,
RA Bevan M., Bancroft I.
RT *Conservation of microstructure between a sequenced region of the
RT genome of rice and multiple segments of the genome of Arabidopsis
RT thaliana.*;
RL Genome Res. 11:1167-1174(2001).
DR EMBL; AJ307662; CAC39074.1; -
KW Hypothetical protein.
SQ SEQUENCE 1012 AA; 108951 MW; 4E8CE3C788BDC44D CRC64;

Alignment Scores:
Pred. No.: 0.653 Length: 1012
Score: 98.50 Matches: 54
Percent Similarity: 34.33% Conservative: 15
Best Local Similarity: 26.87% Mismatches: 84
Query Match: 7.14% Indels: 48
DB: 10 Gaps: 8

US-09-658-824-808 (1-781) x Q949B8 (1-1012)
QY 123 AGAGAACACAGCGCTCCGAGGGTGTCTGTCAGTCAGTCAGAGTCAGAGAGGCCCT 182
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 ArgGlyGluInSerIleProArgGlyGlyArgAlaSerGlySerArgAspArgGlyPro 340
QY 183 CGAAGTCGTCTGCTCTCATGCG- - - - -GTCCCA 212
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 341 GlySerSerArgProAlaAspAlaArgGlyLysArgLysGlnGlyThrProProPro 360
QY 213 CGCCATGGACCTCTCTGTCAGCGCCATACTAGGAGGAAGGCGGAGGAG 272
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 SerPro- - - - -ProArgGlyGly- - - - -GlyAlaAlaArgAlaAsnSer 373
QY 273 TGGAGGGGCTCAGCGAAGCTGGGTGCTGTGGGGTATCCGAGTCCAGAGCACCTG 332
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 ArgArgProGluGlyAlaAlaProThr- - - - - 382
QY 333 GAACCCCGACAGAAGATTCTGGACTCCCGACGGGACGAGAGGAGCGGCATGAGCG 392
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 - - - - -SerGlnProGluGlyLysLysLysLysLysLysLysLysLysLysLys 395
QY 393 ACACACAAACACAGACACAC- - - - -AGCCAGTCCCGAGGAGCCCGAGT 437
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 396 LysMetIlyGluThrGluProSerArgGlyAsnLeuIleSerProLysTrpSerPhe 415
QY 438 AATGGAGAGCCCAAAAGAAAGAACACAGCAGCTGAAAGTCGGGATCCTACACCTGGGCG 497
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 416 AsnArgThrProArgSerGluIleProSerArgProSerArgHisSerLysSerGlyGln 435
QY 498 CAGCACAGAAGATCAGGTATACAGTCAGATCCCATGCGGACATGGAGGTGATCTG 557
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 436 SerGluAlaGluAspProAlaAlaAlaGlu- - - - -AlaArgArgArgGluSerAspArg 453
QY 558 CAAGAG- - - - -CTGCATCAGTCAACACCGGGGATAAATCTGGATTGGGTTC 605
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 ArgGluAlaAlaAspArgLeuArgGluAlaGluAlaGluAlaGluAlaAlaAlaArg 472
QY 606 CGCGCTCAAGGTGAAGATATACCTAAAGAGGAACACTGTAAATGCCAGAGCAGGTGA 665
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 473 AlaArgGlnAlaGluGlu-IleAlaArgGluGluAlaAlaAlaArgAlaArgGlnAlaGlu 492
QY 666 A 666
Db 492 u 492

RESULT 36
Q95KU4
```

```
ID Q95KU4 PRELIMINARY; PRT: 1422 AA.
AC Q95KU4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TREACLE.
GN TCOF1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Haworth K.E., Islam I., Breen M., Putt W., Binns M., Hopkinson D.,
RA Edwards Y.
RT *Canine TCOF1: Cloning, Chromosome Assignment and Genetic analysis in
RT dogs with different face and head types.*;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ296287; CAC82741.1; -
SQ SEQUENCE 1422 AA; 145804 MW; B7381D3EDAF79132 CRC64;

Alignment Scores:
Pred. No.: 0.688 Length: 1422
Score: 98.50 Matches: 55
Percent Similarity: 40.10% Conservative: 22
Best Local Similarity: 28.65% Mismatches: 74
Query Match: 7.14% Indels: 41
DB: 6 Gaps: 9

US-09-658-824-808 (1-781) x Q95KU4 (1-1422)
QY 19 GCGAGTCGGTCCCTGAGGTCCTGGATTCTTCGCTACTGAGACACGGCGGTAGGT 78
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 245 GlyLysAlaGlyPro- - - - -SerAlaAlaGlnAlaLysThrGluLys 258
QY 79 CCACAGCGACATCCAACTGGGAGTGTGAAGTGTGAGTGAGTGAAGAGAACACGAGGC 138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 259 ProLysLysAspSerAspSerSerGluLysAspSerSerGluLysGluProProAla 278
QY 139 TTCGGAGGTTGTGTGGTCACTGAGTCAGTGAGAGGCCCTCGAAGTCGTCGCCCT 198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 AlaLysThrProLeu- - -GlnValLysProSerGlyLysThrProGlnValLysAlaAla 297
QY 199 CTCATCGGTGCCCGCCATCGGCTTCTGCTCGTCAGGCCCATTAAGTGGAGGAA 258
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 SerAlaSerAla- - -LysGluSerProArgLysGlyValProProValProGlyLys 316
QY 259 GGAGGCGCGAGGAGTGAGGGGCTCAGGGCAAGCTGGGTGCTGTTGGGGGTATCCGAGT 318
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 ValGlyProAlaAlaGly- - -GlnAlaLysLys-GlyAla- - - - -Gln 329
QY 319 CCCAGAGCACCTGGAAACCCCGACAGAAGATTCTGACTCCCGAGAC- - - - - 365
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 329 yGluGluAspProAspSerSerThrGluLysSer-AspSerGluGluGluAlaProThrA 349
QY 366 - - - - -GGGACCGAGGAGGCGGATGAGCGGACACACACACACAGAACACACACA 417
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 349 lavalProProThrArg- - - - -SerProValGlnAlaLysProSerG 363
QY 418 GCCAGTCCCGAGGAG- - - - -CCAGTAATGAGAGAGCCCAAGAAAGAG 459
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 yGlnAsnSerGlnValArgThrAlaSerGlyProValLysGlyProGlnLysAlaG 383
QY 460 ACCAGCACCTCAAGTCCGGATCCCTACCTGGGCGAGCAGACAGAAGATCAGGATA 519
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 yProAlaAla- - - - -ThrProValGlyLysGlnGluGluAspSerGluS 398
QY 520 CAGCTGAGATCCAGTGCAGCGACATGGAAGGT 551
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 398 erSerSerGluGluSerAspSerSerGluGly 408

RESULT 37
Q13059
```



